

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 30, 2004, 11:16:19 : Search time 9275 Seconds
(without alignments)
11164.044 Million cell updates/sec
Title: US-09-765-271-55
Perfect score: 2389
Sequence: 1 TTCTTACAGTGGACTGT.....TAAGTAAGAAATAAAC 2389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 90 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sv.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2388	100.0	2389	6	AR120265	AR120265 Sequence
2	2388	100.0	2389	6	AR340956	AR340956 Sequence
3	2388	100.0	2389	6	BD063274	BD063274 Streptoco
4	2388	100.0	2406	6	AX569139	AX569139 Sequence
5	2388	100.0	2451	6	BD268048	BD268048 Streptoco
6	2388	100.0	2451	6	AR344447	AR344447 Sequence
7	2388	100.0	2541	1	AR291695	AR291695 Streptoco
8	2388	100.0	8195	6	AR218862	AR218862 Sequence
9	2388	100.0	8195	6	BD003774	BD003774 Polynucle
10	2388	100.0	10256	1	AE007418	AE007418 Streptoco
11	2388	100.0	349980	6	AX571763	AX571763 Sequence
12	2365.6	99.0	10320	1	AE008479	AE008479 Streptoco
13	2328.2	97.5	20035	2	SPNEU1915	AL449937 Streptoco
14	1434.4	60.0	2535	1	AF340221	AF340221 Streptoco
15	1374.2	57.5	2523	6	BD263585	BD263585 Novel str
16	1374.2	57.5	2523	6	AX343072	AX343072 Sequence
17	1374.2	57.5	2647	6	BD263590	BD263590 Novel str
18	1374.2	57.5	2647	6	AX343073	AX343073 Sequence
19	1011.8	42.4	2478	6	BD229972	BD229972 Human com
20	1003.8	42.0	2481	6	BD264091	BD264091 Streptoco
21	993	41.6	2457	6	AX569137	AX569137 Sequence
22	991	41.5	2457	1	AF318954	AF318954 Streptoco
23	991	41.5	2531	6	BD268049	BD268049 Sequence
24	991	41.5	2531	6	AR344448	AR344448 Sequence
25	990	41.4	2517	1	AF318955	AF318955 Streptoco
26	990	41.4	2517	6	AX568778	AX568778 Sequence
27	990	41.4	2531	6	BD268046	BD268046 Streptoco
28	990	41.4	2531	6	AR344445	AR344445 Sequence
29	990	41.4	11931	1	AE007403	AE007403 Streptoco
30	990	41.4	349980	6	AX571762	AX571762 Sequence
31	988.4	41.4	232807	2	SPNEU1901	AL449923 Streptoco
32	987.6	41.3	2290	6	AR120270	AR120270 Sequence
33	987.6	41.3	2290	6	AR340961	AR340961 Sequence
34	987.6	41.3	2290	6	BD063279	BD063279 Streptoco
35	980.8	41.1	2639	6	BD263591	BD263591 Novel str
36	980.8	41.1	2639	6	AX343074	AX343074 Sequence
37	962.8	40.3	12372	1	AE008464	AE008464 Streptoco
38	869.4	36.4	2166	12	AF340222	AF340222 Synthetic
39	835.8	35.0	75874	2	SPNEU1907	AL449929 Streptoco
40	674	28.2	702	2	SPNEU1929	AL449951 Streptoco
41	653.8	27.4	2359	6	AR219011	AR219011 Sequence
42	653.8	27.4	2359	6	BD003923	BD003923 Polynucle
43	548.4	23.0	3171	6	BD263631	BD263631 Novel str
44	385.4	16.1	1342	6	AR120328	AR120328 Sequence
45	385.4	16.1	1342	6	AR341019	AR341019 Sequence
46	385.4	16.1	1342	6	BD063337	BD063337 Streptoco
47	385.4	16.1	1455	6	BD263587	BD263587 Novel str
48	385.4	16.1	1455	6	BD264131	BD264131 Streptoco
49	385.4	16.1	1455	6	BD268047	BD268047 Streptoco
50	385.4	16.1	1455	6	AR344446	AR344446 Sequence
51	385.4	16.1	1455	6	BD223895	BD223895 Nucleic a
52	385.4	16.1	3117	6	AX568780	AX568780 Sequence
53	385.4	16.1	3120	1	AF318956	AF318956 Streptoco
54	385.4	16.1	3120	6	BD263584	BD263584 Novel str
55	385.4	16.1	3120	6	AX343070	AX343070 Sequence
56	385.4	16.1	5048	6	BD263589	BD263589 Novel str
57	385.4	16.1	5048	6	AX343071	AX343071 Sequence
58	385.4	16.1	6867	6	AR218960	AR218960 Sequence
59	385.4	16.1	6867	6	BD003872	BD003872 Polynucle
60	381.4	16.0	973	6	AR219123	AR219123 Sequence
61	381.4	16.0	973	6	BD004035	BD004035 Polynucle
62	347	14.5	2528	6	BD263592	BD263592 Novel str
63	347	14.5	2528	6	AX343078	AX343078 Sequence
64	253.2	10.6	10312	1	AE007404	AE007404 Streptoco
65	247.4	10.4	504	6	BD229970	BD229970 Human com

```

c 66 247.4 10.4 504 6 BD229971 Human com
67 243.4 10.2 1684 6 AR219026 Sequence
68 243.4 10.2 1684 6 BD003938 Polynucle
c 69 219.6 9.2 75248 6 AX602204 Sequence
c 70 219.6 9.2 98050 1 SAG766854
71 214.6 9.0 2478 6 AX088372 Streptoco
c 72 214.6 9.0 10825 1 AE006623 Streptoco
73 213 8.9 2478 1 AB073859 Streptoco
c 74 213 8.9 10618 1 AE010110 Streptoco
75 211.4 8.8 1146 6 BD263932 Nucleic a
76 211.4 8.8 1910 1 SAG290952 Streptoco
77 211.4 8.8 2469 6 BD263929 Nucleic a
78 211.4 8.8 2469 6 AX088376 Sequence
79 211.4 8.8 2469 6 AX607520 Sequence
80 211.4 8.8 3501 1 AF062533 Streptoco
c 81 211.4 8.8 11617 6 AX602163 Sequence
82 211.4 8.8 21414 1 AE014248 Streptoco
c 83 211.4 8.8 199050 1 SAG766850 Streptoco
84 209.8 8.8 2469 6 BD263632 Novel str
85 209.8 8.8 5215 6 BD136831 Group B s
86 206.6 8.6 2472 6 BD263633 Novel str
c 87 206 8.6 21397 1 AE014279 Streptoco
c 88 205 8.6 50354 1 AE014169 Streptoco
c 89 205 8.6 323825 1 AP005146 Streptoco
90 201.6 8.4 1446 6 AX608407 Sequence

```

ALIGNMENTS

```

RESULT 1
AR120265 LOCUS AR120265 2389 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 55 from patent US 6159469.
ACCESSION AR120265
VERSION AR120265.1 GI:14103841
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 2389)
Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B.,
Fannon,M.R. and Rosen,C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 55 12-DEC-2000;
FEATURES Location/Qualifiers
Source 1..2389
/mol_type="unknown"
/organism="unknown"

Query Match 100.0%; Score 2388; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60

QY 61 TATAGATGAAACACAGCAGCAGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA 120
Db 61 TATAGATGAAACACAGCAGCAGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA 120

QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATAGATAACAGACCAAGGGCTATGTCAC 180
Db 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATAGATAACAGACCAAGGGCTATGTCAC 180

QY 181 TTCATGCGGACCACTATCAATTAATGTAAGGTTCCCTATGACGCTATCATCAG 240
Db 181 TTCATGCGGACCACTATCAATTAATGTAAGGTTCCCTATGACGCTATCATCAG 240

QY 241 TGAAGAATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGGATATGTTAATCA 300

```

```

241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAATGA 300
301 GGTCAAGGGTGGATATGTTATCAAGCTAGATGAAATACTATGTTTACCTTAAGATGC 360
301 GGTCAAGGGTGGATATGTTATCAAGCTAGATGAAATACTATGTTTACCTTAAGATGC 360
361 TGCCCAACGGGATAACGTCGGTACAAAGAGGAAAATCAATCGACAAAACCAAGAGATAG 420
361 TGCCCAACGGGATAACGTCGGTACAAAGAGGAAAATCAATCGACAAAACCAAGAGATAG 420
421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCCTGGCAGCTTCGA 480
421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCCTGGCAGCTTCGA 480
481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
541 TGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTAATTCCTAAGAAATGAGTT 600
541 TGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTAATTCCTAAGAAATGAGTT 600
601 ATCAGTAGCGAGTTGGCTGCTGAGAACGCTTCTATCTGTCGAGGAAATCTGTCAA 660
601 ATCAGTAGCGAGTTGGCTGCTGAGAACGCTTCTATCTGTCGAGGAAATCTGTCAA 660
661 TTCAGAACCTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTACTCTTC 720
661 TTCAGAACCTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTACTCTTC 720
721 TGTAGCAATCCAGGAATCTACAAATACTACAAACAAAGCAACAACTAACAGTCA 780
721 TGTAGCAATCCAGGAATCTACAAATACTACAAACAAAGCAACAACTAACAGTCA 780
781 AGCAAGTCAAGATATGACATTTAGTCTCTTTGAACAGTCTTACAACTGCTTTGAG 840
781 AGCAAGTCAAGATATGACATTTAGTCTCTTTGAACAGTCTTACAACTGCTTTGAG 840
841 TCAACGACATGAGAAATCTGATGGCTTCTTTGATCCAGCACAATCACAAGTCGAAC 900
841 TCAACGACATGAGAAATCTGATGGCTTCTTTGATCCAGCACAATCACAAGTCGAAC 900
901 AGCTAGAGTGTGTCAGTGCACACGGAGATCATTAACCACTTCACTCTTACTCTCAAT 960
901 AGCTAGAGTGTGTCAGTGCACACGGAGATCATTAACCACTTCACTCTTACTCTCAAT 960
961 GTCTGAATTTGGAAGAACGAAATCGCTGATATTTCCCTTCTGTTATCGTTCAACATG 1020
961 GTCTGAATTTGGAAGAACGAAATCGCTGATATTTCCCTTCTGTTATCGTTCAACATG 1020
1021 GGTACCAGATTCAAGGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGS 1080
1021 GGTACCAGATTCAAGGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGS 1080
1081 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAATCTCTTTGGTTAGTCAGCTGT 1140
1081 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAATCTCTTTGGTTAGTCAGCTGT 1140
1141 AGCAAGTGTGGGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTTC 1200
1141 AGCAAGTGTGGGAGGATATGTTATCGAAGAAAGGCACTCTCTGTTATGCTTTTC 1200
1201 GAAAGATTTACCCTGTAACCTGTTAAATAATCTTGAAGCAAGTATCAAAAACAGAGAG 1260
1201 GAAAGATTTACCCTGTAACCTGTTAAATAATCTTGAAGCAAGTATCAAAAACAGAGAG 1260
1261 TGTTTTCAACACTTAACTGCTTAAATAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
1261 TGTTTTCAACACTTAACTGCTTAAATAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
1321 TCAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGGTCG 1380
1321 TCAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGGTCG 1380

```

QY	1381	TAATTTCTGATTTCCAGCCTTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA	1444		
Db	1381	TAATTTCTGATTTCCAGCCTTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA	1440		
QY	1441	TAAAGAAAATTTGGTAGATGATTTATTGGCATTCCTTAGACCAATTAACCCATCCAGAGCG	1500		
Db	1441	TAAAGAAAATTTGGTAGATGATTTATTGGCATTCCTTAGACCAATTAACCCATCCAGAGCG	1500		
QY	1501	ACTTGGCAAAACCAAAATTCCTCAAATTTAGTATACTGAAGACGAAGTTTCGTATTGCTCAATTT	1560		
Db	1501	ACTTGGCAAAACCAAAATTCCTCAAATTTAGTATACTGAAGACGAAGTTTCGTATTGCTCAATTT	1560		
QY	1561	AGCTGATAAGTATACAAACGTCAGATGCTTACATTTTGTGTAACATGATATAATCACTGTA	1620		
Db	1561	AGCTGATAAGTATACAAACGTCAGATGCTTACATTTTGTGTAACATGATATAATCACTGTA	1620		
QY	1621	TGAAGGAGATGCATATGTAAGCCTCATATGGGCCATAGTCACTGGATTCGAAAAAGATAG	1680		
Db	1621	TGAAGGAGATGCATATGTAAGCCTCATATGGGCCATAGTCACTGGATTCGAAAAAGATAG	1680		
QY	1681	CTTTCTTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGTATCCTTACC	1740		
Db	1681	CTTTCTTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGTATCCTTACC	1740		
QY	1741	TCCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCAGCAGCTATTTA	1800		
Db	1741	TCCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCAGCAGCTATTTA	1800		
QY	1801	CAATCGTGTGAAAGGGGAAAAAGAAATTCACCTCGTTCGACTTCCTCATATATGTTGAGCA	1860		
Db	1801	CAATCGTGTGAAAGGGGAAAAAGAAATTCACCTCGTTCGACTTCCTCATATATGTTGAGCA	1860		
QY	1861	TACAGTTTGAGTTTAAAAACGGTAATTTGATTATTTCCTCATAAAGGATCATTTACCATATAT	1920		
Db	1861	TACAGTTTGAGTTTAAAAACGGTAATTTGATTATTTCCTCATAAAGGATCATTTACCATATAT	1920		
QY	1921	TAAATTTGCTTGGTTTGATGATCACACATACAAGCTCCAAATGGCTATACCTTGGGAAGA	1980		
Db	1921	TAAATTTGCTTGGTTTGATGATCACACATACAAGCTCCAAATGGCTATACCTTGGGAAGA	1980		
QY	1981	TTTGTGTTGCGACGANTAAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCAATGA	2040		
Db	1981	TTTGTGTTGCGACGANTAAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCAATGA	2040		
QY	2041	TGGATGGGGCAATGCCAGTCGAGATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	2100		
Db	2041	TGGATGGGGCAATGCCAGTCGAGATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	2100		
QY	2101	TAAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCCC	2160		
Db	2101	TAAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCCC	2160		
QY	2161	TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGAGAAAGTTTGTGTCGCAA	2220		
Db	2161	TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGAGAAAGTTTGTGTCGCAA	2220		
QY	2221	AGTAAACGGAATTCAGTCTGAAAGCCATGCAACAGAAACCTAGCTGTTTACGAATAA	2280		
Db	2221	AGTAAACGGAATTCAGTCTGAAAGCCATGCAACAGAAACCTAGCTGTTTACGAATAA	2280		
QY	2281	TTTGACTCTTCAAATTTGATTAACAATAGTATCATGGCAAGACGAAAAAATTTACTTGC	2340		
Db	2281	TTTGACTCTTCAAATTTGATTAACAATAGTATCATGGCAAGACGAAAAAATTTACTTGC	2340		
QY	2341	GTGTTTAAAAAGGAGTAAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC	2389		
Db	2341	GTGTTTAAAAAGGAGTAAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC	2389		
RESULT 2	AR340956	2389 bp	DNA	linear	PAT 17-AUG-2003

RESULT 2
AR340956
LOCUS

QY 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 840
 Db 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 840
 QY 841 TCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGACAAATCACAGTCAAC 900
 Db 841 TCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGACAAATCACAGTCAAC 900
 QY 901 AGCTAGAGTGTGCAAGTGCACACGAGATCAATACCACTTCAATCTCTCAAT 960
 Db 901 AGCTAGAGTGTGCAAGTGCACACGAGATCAATACCACTTCAATCTCTCAAT 960
 QY 961 GTCTGAATTTGGAAGAACGAAATCGCTCGTATTTATTCCTTCCCTTTCGTTATCGTTCAAAACATTG 1020
 Db 961 GTCTGAATTTGGAAGAACGAAATCGCTCGTATTTATTCCTTCCCTTTCGTTATCGTTCAAAACATTG 1020
 QY 1021 GGTACAGATTTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTTAGTCCAG 1080
 Db 1021 GGTACAGATTTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTTAGTCCAG 1080
 QY 1081 CCGCAACCTGCACCAAACTTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
 Db 1081 CCGCAACCTGCACCAAACTTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
 QY 1141 ACGAAAAGTTGGGGAAGGATATGATTTCAAGAAAAAGGGCATCTCTCGTTATGCTTTTC 1200
 Db 1141 ACGAAAAGTTGGGGAAGGATATGATTTCAAGAAAAAGGGCATCTCTCGTTATGCTTTTC 1200
 QY 1201 GAAAGATTTACCAATCTGAACTGTTAAATCTTTGAAAGCAAGTTATCAAAACAGAGAG 1260
 Db 1201 GAAAGATTTACCAATCTGAACTGTTAAATCTTTGAAAGCAAGTTATCAAAACAGAGAG 1260
 QY 1261 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCGCTGACCAAAATTTTA 1320
 Db 1261 TGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCGCTGACCAAAATTTTA 1320
 QY 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGGCTTGTGTTGNAATAAAGGCTG 1380
 Db 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGGCTTGTGTTGNAATAAAGGCTG 1380
 QY 1381 TAATTCGTATTTCCAGCCTTAGACAAATTTAGAAAGCGTTGAATGATGATCGACTAA 1440
 Db 1381 TAATTCGTATTTCCAGCCTTAGACAAATTTAGAAAGCGTTGAATGATGATCGACTAA 1440
 QY 1441 TAAAGAAAAATGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGAG 1500
 Db 1441 TAAAGAAAAATGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGAG 1500
 QY 1501 ACTTGGCAAAACCAAAATTTCTCAATTTAGTATATACTGAAGACGAAAGTTTCGTTTCTCAAT 1560
 Db 1501 ACTTGGCAAAACCAAAATTTCTCAATTTAGTATATACTGAAGACGAAAGTTTCGTTTCTCAAT 1560
 QY 1561 AGCTGATAGTATACAAAGTCAAGTGGTATGATTTGATGAACATGATATACAGTGA 1620
 Db 1561 AGCTGATAGTATACAAAGTCAAGTGGTATGATTTGATGAACATGATATACAGTGA 1620
 QY 1621 TGAAGGAGATGCATATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
 Db 1621 TGAAGGAGATGCATATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680
 QY 1681 CTTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTTAAAGAAAAAGGTTATCTTACC 1740
 Db 1681 CTTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTTAAAGAAAAAGGTTATCTTACC 1740
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCAACTGGAGTAGTGCAGCTATTATTA 1800
 Db 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCAACTGGAGTAGTGCAGCTATTATTA 1800
 QY 1801 CAACTCGTGAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCCATATATGTTGAGCA 1860
 Db 1801 CAACTCGTGAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCCCATATATGTTGAGCA 1860

QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCAATTACCATAATAT 1920
 Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCAATTACCATAATAT 1920
 QY 1921 TAAATTTGCTTGGTTTGTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
 Db 1921 TAAATTTGCTTGGTTTGTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
 QY 1981 TTTGTTTGGGAGGATTAAGTACTAGTAGAACACCTGACGACGCTCCACATTTCTAATGA 2040
 Db 1981 TTTGTTTGGGAGGATTAAGTACTAGTAGAACACCTGACGACGCTCCACATTTCTAATGA 2040
 QY 2041 TGGATGGGCAATGCGCAGTGGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 Db 2041 TGGATGGGCAATGCGCAGTGGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 QY 2101 TGAAGATTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC 2160
 Db 2101 TGAAGATTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCC 2160
 QY 2161 TCAAGTAGAGACTTGAAGAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTTCCTTCGGA 2220
 Db 2161 TCAAGTAGAGACTTGAAGAAAGTAGAGCCCAACTCAAGAAAGCAGAAAGTTTTCCTTCGGA 2220
 QY 2221 AGTAACGGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATA 2280
 Db 2221 AGTAACGGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATA 2280
 QY 2281 TTTGACTTTCAAAATTTATGATTAACATAGTATCATGCGAGAGACGAAATAATCTTTC 2340
 Db 2281 TTTGACTTTCAAAATTTATGATTAACATAGTATCATGCGAGAGACGAAATAATCTTTC 2340
 QY 2341 GTTGTAAAAAGGAAGTAATCTTCTATCTGTAAGTAAAGAAAAATAAAC 2389
 Db 2341 GTTGTAAAAAGGAAGTAATCTTCTATCTGTAAGTAAAGAAAAATAAAC 2389

RESULT 3
 BD063274 2389 bp DNA linear PAT 27-AUG-2002
 LOCUS Streptococcus pneumoniae antigens and vaccines.
 DEFINITION BD063274
 ACCESSION BD063274.1 GI:22608877
 VERSION JP 2001505415-A/28
 KEYWORDS unclassified
 SOURCE unclassified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2389)
 AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromockyj,A.
 TITLE Streptococcus pneumoniae antigens and vaccines
 JOURNAL Patent: JP 2001505415-A 28 24-APR-2001;
 HUMAN GENOME SCIENCES INC
 COMMENT PN JP 2001505415-A/28
 PD 24-APR-2001
 PF 30-OCT-1997 JP 1998520667
 PR 31-OCT-1996 US 60/029960
 PI CHARLES A KUNSCH,GIL H CHOI,SYDNOR L JOHNSON,ALEX HROMOCKYJ PC
 C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
 G01N33/569,
 PC GOIN33/68
 CC Strandedness: Double;
 CC Topology: Linear;
 PH Key Location/Qualifiers.
 FEATURES
 1. 2389
 /organism="unclassified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 ORIGIN
 Query Match 100.0%; Score 2388; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTGTTCCTA 60
 Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTGTTCCTA 60
 QY 61 TATAGATGGAAAAAAGCGACGCAAAACGAGAAATTTGACTCTCGATGAGGTTAGCAA 120
 Db 61 TATAGATGGAAAAAAGCGACGCAAAACGAGAAATTTGACTCTCGATGAGGTTAGCAA 120
 QY 121 GCGTGAAGGAATCAATGTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCCAC 180
 Db 121 GCGTGAAGGAATCAATGTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCCAC 180
 QY 181 TTCACATGCGACCACTATCATTTATCAAGTGAAGGTTCTTATGACGCTATCATCAG 240
 Db 181 TTCACATGCGACCACTATCATTTATCAAGTGAAGGTTCTTATGACGCTATCATCAG 240
 QY 241 TGAAGAAATTAATCTATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTTGTTAATGA 300
 Db 241 TGAAGAAATTAATCTATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTTGTTAATGA 300
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTTAAGGATGC 360
 Db 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTTAAGGATGC 360
 QY 361 TGCCACGCGGATACGTCGTTACAAAAGAGAAATCAATCGACAAAACAGAGCATAG 420
 Db 361 TGCCACGCGGATACGTCGTTACAAAAGAGAAATCAATCGACAAAACAGAGCATAG 420
 QY 421 TCAACATCGTGAAGGTGAAACTCCAAGAAACGATGGTGTGTGCTTTGGCACGTTGCA 480
 Db 421 TCAACATCGTGAAGGTGAAACTCCAAGAAACGATGGTGTGTGCTTTGGCACGTTGCA 480
 QY 481 AGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 Db 481 AGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 QY 541 TGGTGATGCTTATATCGTTTCTCATGGAGATCATTTACCATACATTTCTTAAGATGAGTT 600
 Db 541 TGGTGATGCTTATATCGTTTCTCATGGAGATCATTTACCATACATTTCTTAAGATGAGTT 600
 QY 601 ATCAGCTAGCGAGTTGGCTGTGCAAGAACGCTTCTATCTGTGCGAGGAAATCTGTCAA 660
 Db 601 ATCAGCTAGCGAGTTGGCTGTGCAAGAACGCTTCTATCTGTGCGAGGAAATCTGTCAA 660
 QY 661 TTCAGAACCTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAAAACGCGTACCTTC 720
 Db 661 TTCAGAACCTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAAAACGCGTACCTTC 720
 QY 721 TGTAAAGCAATCCAGGAATCAAAATTAACAACAGCAACACAGCAACACTTAACAGTCA 780
 Db 721 TGTAAAGCAATCCAGGAATCAAAATTAACAACAGCAACACAGCAACACTTAACAGTCA 780
 QY 781 AGCAAGTCAAAAGTAATGATGATGATGCTCTTTGAAACAGCTCTCAAAACGCTTTGAG 840
 Db 781 AGCAAGTCAAAAGTAATGATGATGATGCTCTTTGAAACAGCTCTCAAAACGCTTTGAG 840
 QY 841 TCAAGCATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAAAGTCGAAC 900
 Db 841 TCAAGCATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAAAGTCGAAC 900
 QY 901 AGCTAGAGGTGTGAGTGGCCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
 Db 901 AGCTAGAGGTGTGAGTGGCCACAGGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
 QY 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTCTGTTATCGTTCAAAACATTG 1020
 Db 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTCTGTTATCGTTCAAAACATTG 1020
 QY 1021 GGTACCAATTTCAAGGCGAGAACCAACAGTCCCAACCGGACTCCGGAACTTAGTCCAGG 1080
 Db 1021 GGTACCAATTTCAAGGCGAGAACCAACAGTCCCAACCGGACTCCGGAACTTAGTCCAGG 1080

QY 1081 CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1140
 Db 1081 CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT 1140
 QY 1141 ACGAAAAGTTGGGGAAGGATATGATTCGAAGAAAAGGCGATCTCTCGTTATGTCTTTGC 1200
 Db 1141 ACGAAAAGTTGGGGAAGGATATGATTCGAAGAAAAGGCGATCTCTCGTTATGTCTTTGC 1200
 QY 1201 GAAAGATTTACCATCTGAAAACCTGTTAAAAATCTTGAAGAGCAAGTTATCAAAACAAGAG 1260
 Db 1201 GAAAGATTTACCATCTGAAAACCTGTTAAAAATCTTGAAGAGCAAGTTATCAAAACAAGAG 1260
 QY 1261 TGTTCACACACTTTAACTGCTAAAAAGAAAATGCTCTCTCTGCTGACCAAGAAATTTA 1320
 Db 1261 TGTTCACACACTTTAACTGCTAAAAAGAAAATGCTCTCTCTGCTGACCAAGAAATTTA 1320
 QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380
 Db 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380
 QY 1381 TAAATCTGATTTCCAAAGCCTTAGACAAATTAATTAGAACGCTTGAATGATGATCACTAA 1440
 Db 1381 TAAATCTGATTTCCAAAGCCTTAGACAAATTAATTAGAACGCTTGAATGATGATCACTAA 1440
 QY 1441 TAAAGAAAATTTGGTAGATGATTTATTGCGCATTCCTAGCACCAATTAACCCATCCAGAGC 1500
 Db 1441 TAAAGAAAATTTGGTAGATGATTTATTGCGCATTCCTAGCACCAATTAACCCATCCAGAGC 1500
 QY 1501 ACTTGGCAAAACCAAAATCTCAAAATTTAGTATATCTGAAGACGAAAGTTCTGTTCAAT 1560
 Db 1501 ACTTGGCAAAACCAAAATCTCAAAATTTAGTATATCTGAAGACGAAAGTTCTGTTCAAT 1560
 QY 1561 AGCTGATAAGTATACAACTCAGATGTTTATGATGAACATGATATATCACTGTA 1620
 Db 1561 AGCTGATAAGTATACAACTCAGATGTTTATGATGAACATGATATATCACTGTA 1620
 QY 1621 TGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 Db 1621 TGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 QY 1681 CCTTTCGATGAAGAAAAGTTGAGGCTCAAGCCTTACTAAAGAAAAGGTTATCTTACC 1740
 Db 1681 CCTTTCGATGAAGAAAAGTTGAGGCTCAAGCCTTACTAAAGAAAAGGTTATCTTACC 1740
 QY 1741 TCCATCTCAGACGAGATGTTTAAAGGAAATCCTGAGATGATGAGAGCTATTTA 1800
 Db 1741 TCCATCTCAGACGAGATGTTTAAAGGAAATCCTGAGATGATGAGAGCTATTTA 1800
 QY 1801 CAATCGTGTGAAGGGGAAAACGAAATTCCTGCTTCCATATATGTTGAGCA 1860
 Db 1801 CAATCGTGTGAAGGGGAAAACGAAATTCCTGCTTCCATATATGTTGAGCA 1860
 QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCAATAT 1920
 Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCAATAT 1920
 QY 1921 TAAATTTGCTTGGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 Db 1921 TAAATTTGCTTGGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 QY 1981 TTTGTTGCGACGATTAAGTACTAGTAGAACACCTCGACCAACCTCCACATTTCTAATGA 2040
 Db 1981 TTTGTTGCGACGATTAAGTACTAGTAGAACACCTCGACCAACCTCCACATTTCTAATGA 2040
 QY 2041 TGGATGGGGCAATGCCAGTGAGCATGTTTGAAGCAAGAAACACAGTGAAGATCCAAA 2100
 Db 2041 TGGATGGGGCAATGCCAGTGAGCATGTTTGAAGCAAGAAACACAGTGAAGATCCAAA 2100
 QY 2101 TAAAGACTTCAAGCGGATGAAGAGCCAGTGAAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
 Db 2101 TAAAGACTTCAAGCGGATGAAGAGCCAGTGAAGAGAAACACCTGCTGAGCCAGAGTCCC 2160
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGACGAGAGTTTGTGTCGAA 2220

1561 AGCTGATAGTATCAACGTCAGATGGTTACATTTTGTGTAACATGATATTAATCAGTGA 1620
 1578 AGCTGATAGTATCAACGTCAGATGGTTACATTTTGTGTAACATGATATTAATCAGTGA 1637
 1621 TGAAGGAGATGATATGTAACGCTCATATGGCCATAGTCACTGGATGGAAAAAGATAG 1680
 1638 TGAAGGAGATGATATGTAACGCTCATATGGCCATAGTCACTGGATGGAAAAAGATAG 1697
 1681 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCTTATATCTAAAGAAAAAGGTTATCTTACC 1740
 1698 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCTTATATCTAAAGAAAAAGGTTATCTTACC 1757
 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1800
 1758 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1817
 1801 CAATCGTGTGAAGGGGAAAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1860
 1818 CAATCGTGTGAAGGGGAAAAAGCAATCCAACTGGAGTACTGCGAGCTATTATTA 1877
 1861 TACAGTTGAGTTAAAGCGGTAATTTGATTTTCTCTATAGGATCAATACCAATATAT 1920
 1878 TACAGTTGAGTTAAAGCGGTAATTTGATTTTCTCTATAGGATCAATACCAATATAT 1937
 1921 TAAATTTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGCTATACCTTGGGAAGA 1980
 1938 TAAATTTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGCTATACCTTGGGAAGA 1997
 1981 TTTGTTGCGAGTAAGTACTAGTAGAAACACCTCTGACGAACGTCACATCTTAATGA 2040
 1998 TTTGTTGCGAGTAAGTACTAGTAGAAACACCTCTGACGAACGTCACATCTTAATGA 2057
 2041 TGGATGGGCAATGCGAGTACGATGTTAGGCAAGAAAGCAACAGTGAAGATCCAAA 2100
 2058 TGGATGGGCAATGCGAGTACGATGTTAGGCAAGAAAGCAACAGTGAAGATCCAAA 2117
 2101 TAAAGAACTTCAAAGCGGATGAAGAGCCAGTACAGGAAACACCTCTGAGCCAGAAAGTCCC 2160
 2118 TAAAGAACTTCAAAGCGGATGAAGAGCCAGTACAGGAAACACCTCTGAGCCAGAAAGTCCC 2177
 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAAAGCAGAAAGTTTTCCTTGGCAA 2220
 2178 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAAAGCAGAAAGTTTTCCTTGGCAA 2237
 2221 AGTAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
 2238 AGTAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2297
 2281 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAGCAGAAAAATTAATCTTGC 2340
 2298 TTTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAGAGCAGAAAAATTAATCTTGC 2357
 2341 GTTGTAAAGGAAAGTAATCTTCTATCTGTAGTAAGGAAAAATAAATTAAC 2389
 2358 GTTGTAAAGGAAAGTAATCTTCTATCTGTAGTAAGGAAAAATAAATTAAC 2406

RESULT 5
 BD268048
 LOCUS
 DEFINITION Streptococcus pneumoniae protein and immunogenic fragments for vaccines.
 ACCESSION BD268048
 VERSION BD268048.1 GI:33077816
 KEYWORDS JP 2002532561-A/6.
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2451)
 AUTHORS Johnson, L.S., Koenig, S. and Adamou, J.E.
 TITLE Streptococcus pneumoniae protein and immunogenic fragments for

JOURNAL Patent JP 2002532561-A 6 02-OCT-2002;
 MEDIMUNE INC
 COMMENT OS Streptococcus pneumoniae
 PN JP 2002532561-A/6
 PD 02-OCT-2002
 PF 21-DEC-1999 JP 2000589215
 PR 21-DEC-1998 US 60/113048
 PI LESLIE, S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU
 PC A61K39/09, A61K38/00, A61P31/10, A61K37/02
 CC Streptococcus pneumoniae protein and immunogenic fragments for vaccines
 Key Location/Qualifiers
 PH 1..2451
 FT source
 FT Location/Qualifiers
 1..2451
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1313"
 FEATURES
 source
 ORIGIN
 Query Match 100.0%; Score 2388; DB 6; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTTCTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAAAATAATCGTGTTCCTA 60
 Db 60 TTTCTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAAAATAATCGTGTTCCTA 119
 QY 61 TATAGATGAAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 Db 120 TATAGATGAAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 179
 QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCCAC 180
 Db 180 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCCAC 239
 QY 181 TTCATGCGGACCACTATCATATTATTAATGTTAAGGTTCTTTATGACGCTATCATCAG 240
 Db 240 TTCATGCGGACCACTATCATATTATTAATGTTAAGGTTCTTTATGACGCTATCATCAG 299
 QY 241 TGAAGAAATTTCTATGAAGATCCAACTATAAGCTAAAGTAAAGTATGTTAAATGA 300
 Db 300 TGAAGAAATTTCTATGAAGATCCAACTATAAGCTAAAGTAAAGTATGTTAAATGA 359
 QY 301 GGTCAAGGCTGGATATGTTATCAAGGTAGATGGAATAATATCTGTTTACCTTAAAGATGC 360
 Db 360 GGTCAAGGCTGGATATGTTATCAAGGTAGATGGAATAATATCTGTTTACCTTAAAGATGC 419
 QY 361 TGCCACGCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAACAGAGCATAG 420
 Db 420 TGCCACGCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAACAGAGCATAG 479
 QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAAGATGCTGCTGCTTGGCAGCGTTCCCA 480
 Db 480 TCAACATCGTGAAGTGGAACTCCAAAGAAAGATGCTGCTGCTTGGCAGCGTTCCCA 539
 QY 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 Db 540 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 599
 QY 541 TGGTATGCTTATATCTGTTCTCTCATGAGATCATACCATTTACATTCTCTAAGATGAGTT 600
 Db 600 TGGTATGCTTATATCTGTTCTCTCATGAGATCATACCATTTACATTCTCTAAGATGAGTT 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAGCGCTTCTATCTGCTGAGGAAATCTGTCAA 660
 Db 660 ATCAGCTAGCGAGTTGGCTGCTGAGAGCGCTTCTATCTGCTGAGGAAATCTGTCAA 719
 QY 661 TTCAAGAACTTATCGCCGACAAAAATAGCGATAACACCTTCAAGAAACAACTGGTACCTTC 720
 Db 720 TTCAAGAACTTATCGCCGACAAAAATAGCGATAACACCTTCAAGAAACAACTGGTACCTTC 779

Qy	1801	CAATCGTGCGAAGGGGAAAAAGAAATTCACCTCGTGTGCACTTCCATATATATGTTGAGCA	1861
Db	1860	CAATCGTGTCGAAAGGGGAAAAAGAAATTCACCTCGTGTGCACTTCCATATATGTTGAGCA	1919
Qy	1861	TACAGTTGAGCGTTAAAAACCGTAATTTGATTATTTCCTCATAAAGATCATTTACCATATAT	1920
Db	1920	TACAGTTGAGGTTAAAAACCGTAATTTGATTATTTCCTCATAAAGATCATTTACCATATAT	1979
Qy	1921	TAAATTTGCTTGTTGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA	1980
Db	1980	TAAATTTGCTTGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA	2039
Qy	1981	TTTGTGTTGCGACGATTAAGTACTAGCTAGAAACACCCGTGACGAGCTCCACATCTTAATGA	2040
Db	2040	TTTGTGTTGCGACGATTAAGTACTAGCTAGAAACACCCGTGACGAGCTCCACATCTTAATGA	2099
Qy	2041	TGAGTGGGGCAATGCCAGTCGAGCATGTGTTAGCGACAGAAAGACACACAGTCGAAGATCCAA	2100
Db	2100	TGAGTGGGGCAATGCCAGTGAGCATGTGTTAGCGAAGAAAGACACACAGTCGAAGATCCAA	2159
Qy	2101	TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGCAAGATCCC	2160
Db	2160	TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGCAAGATCCC	2219
Qy	2161	TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGACGAAAGTTTGTCTTGGCAA	2220
Db	2220	TCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAAGAGACGAAAGTTTGTCTTGGCAA	2279
Qy	2221	AGTAAACGATTCCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA	2280
Db	2280	AGTAAACGATTCCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA	2339
Qy	2281	TTTGACTCTTCAAATATTGATTAACATAGTATCATGGCAGAAAGCAAAAAATTTACTTGC	2340
Db	2340	TTTGACTCTTCAAATATTGATTAACATAGTATCATGGCAGAAAGCAAAAAATTTACTTGC	2399
Qy	2341	GTTGTTAAAAAGGAGTAATCCTTCATCTGTAAGTAAAGAAAAATAAAC	2389
Db	2400	GTTGTTAAAAAGGAGTAATCCTTCATCTGTAAGTAAAGAAAAATAAAC	2448

RESULT 6
 AR344447 LOCUS
 AR344447 DEFINITION
 AR344447 ACCESSION
 AR344447 VERSION
 AR344447.1 GI:33740486
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unkown.
 REFERENCE Unclassified.
 1 (bases 1 to 2451)
 Johnson L.S. and Adamou, J.E.
 Vaccine compositions comprising Streptococcus pneumoniae
 polypeptides having selected structural MOTIFS
 Patent: US 6582706-A 9 24-JUN-2003;
 Location/Qualifiers
 1..2451
 /organism="unkown"
 /mol type="genomic DNA"
 JOURNAL
 FEATURES source

ORIGIN	/mol_type= genomic DNA				
Query Match	100.0%	Score 2388;	DB 6;	Length 2451;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 2389;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	TTCTTAGAGTTGGAGCTGTATCAAGCTAGACGGTTAAGGAAAATAATCGTGTTCCTA	60		
Db	60	TTCTTAGAGTTGGAGCTGTATCAAGCTAGACGGTTAAGGAAAATAATCGTGTTCCTA	119		
Qy	61	TATAGATGGAAAAACAAGCGACGCAAAAACCGAGAAATTTGACTCTCATGAGGTTAGCAA	120		
Db	120	TATAGATGGAAAAACAAGCGACGCAAAAACCGAGAAATTTGACTCTCATGAGGTTAGCAA	179		

QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
 Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 239
 QY 181 TTCACATGGCAGCACTATCATTTATTAATGTTAAAGGTTCTTATGACGCTATCATCAG 240
 Db 240 TTCACATGGCAGCACTATCATTTATTAATGTTAAAGGTTCTTATGACGCTATCATCAG 299
 QY 241 TGAAGAAATTAATCTGAAGAGTCCAAAATTAAGCTTAAAGATGAGGATATGTTAAATGA 300
 Db 300 TGAAGAAATTAATCTGAAGAGTCCAAAATTAAGCTTAAAGATGAGGATATGTTAAATGA 359
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGTGGAAAAATCTATGTTTACCTTTAAGGATGC 360
 Db 360 GGTCAAGGGTGGATATGTTATCAAGGTAGTGGAAAAATCTATGTTTACCTTTAAGGATGC 419
 QY 361 TGCCACGCGGATAACGTCGCTACAAAAGAGAAATCAATCGACAAAAACAAGACATAG 420
 Db 420 TGCCACGCGGATAACGTCGCTACAAAAGAGAAATCAATCGACAAAAACAAGACATAG 479
 QY 421 TCAACATCGTGAAGGTGAACTCCAAAGAACGATGGTCTGTTGCTTTGGCACGTTCCGA 480
 Db 480 TCAACATCGTGAAGGTGAACTCCAAAGAACGATGGTCTGTTGCTTTGGCACGTTCCGA 539
 QY 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGTAC 540
 Db 540 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGTAC 599
 QY 541 TGGTGATCGTTATATCGTTCTCATGGAGATCATTTACCAATACATTTCTTAAGATGAGTT 600
 Db 600 TGGTGATCGTTATATCGTTCTCATGGAGATCATTTACCAATACATTTCTTAAGATGAGTT 659
 QY 601 ATCAGCTAGAGTGGTGGTGCAGAGAGCCCTTCTATCTGTTGAGGAAATCTGTCAAA 660
 Db 660 ATCAGCTAGAGTGGTGGTGCAGAGAGCCCTTCTATCTGTTGAGGAAATCTGTCAAA 719
 QY 661 TTCAGAACCTATCGCCACAAAATAGCGATTAACACTTCNAGAACAACTGGTACCTTC 720
 Db 720 TTCAGAACCTATCGCCACAAAATAGCGATTAACACTTCNAGAACAACTGGTACCTTC 779
 QY 721 TGTAAAGCAATCAGGAATCAAAATTAACAAAGCAACACAGCAACACATTAACAGTCA 780
 Db 780 TGTAAAGCAATCAGGAATCAAAATTAACAAAGCAACACAGCAACACATTAACAGTCA 839
 QY 781 AGCAAGTCAAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 Db 840 AGCAAGTCAAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 899
 QY 841 TCACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCACAAATCAAGTCCGAAC 900
 Db 900 TCACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCACAAATCAAGTCCGAAC 959
 QY 901 AGCTAGAGGTGTGAGTGCACACGAGATCATTTACCACTTTCATCCCTTACTCTCAAT 960
 Db 960 AGCTAGAGGTGTGAGTGCACACGAGATCATTTACCACTTTCATCCCTTACTCTCAAT 1019
 QY 961 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTAATTCCTTCTGTTATGTTTCAAACTTG 1020
 Db 1020 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTAATTCCTTCTGTTATGTTTCAAACTTG 1079
 QY 1021 GGTACCAATTCAGGCGACAGACACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 Db 1080 GGTACCAATTCAGGCGACAGACACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139
 QY 1081 CCCGCAACTGACCAAAATCTTTAAATAGACTCAAAATCTTCTTTGTTAGTCACTGCT 1140
 Db 1140 CCCGCAACTGACCAAAATCTTTAAATAGACTCAAAATCTTCTTTGTTAGTCACTGCT 1199
 QY 1141 AGGAAAGTGGGGAAGGATGATGATTCGAGAAAGGCGATCTCTGCTTATGCTTTTC 1200
 Db 1200 AGGAAAGTGGGGAAGGATGATGATTCGAGAAAGGCGATCTCTGCTTATGCTTTTC 1259

QY 1201 GAAGATTTACATCTGAACCTGTTTAAATACTTTGAAGCAAGTTTATCAAAACAGAGAG 1260
 Db 1260 GAAGATTTACATCTGAACCTGTTTAAATACTTTGAAGCAAGTTTATCAAAACAGAGAG 1319
 QY 1261 TGTTTACACACACTTTAACTGCTGCTAAAAAAGAAATGTTCTCTCTGCTGCTGCTGCTGCT 1320
 Db 1320 TGTTTACACACACTTTAACTGCTGCTAAAAAAGAAATGTTCTCTCTCTGCTGCTGCTGCTGCT 1379
 QY 1321 TGAATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380
 Db 1380 TGAATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1439
 QY 1381 TAATTCGATTTCCAAGCCTTAGACAAATTAAGACGCTTGAATGATGATGATGATGATGATG 1440
 Db 1440 TAATTCGATTTCCAAGCCTTAGACAAATTAAGACGCTTGAATGATGATGATGATGATGATG 1499
 QY 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTTACCCATCCAGAGCG 1500
 Db 1500 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTTACCCATCCAGAGCG 1559
 QY 1501 ACTTGGGCAAAACCAATTTCTCAAAATTTAGTATATCTGAGACGAAAGTTCTGTTTCTCAAT 1560
 Db 1560 ACTTGGGCAAAACCAATTTCTCAAAATTTAGTATATCTGAGACGAAAGTTCTGTTTCTCAAT 1619
 QY 1561 AGCTGATAAGTATACACGCTCAGATGTTTACATTTTTCATGAACATGATATAATCAGTGA 1620
 Db 1620 AGCTGATAAGTATACACGCTCAGATGTTTACATTTTTCATGAACATGATATAATCAGTGA 1679
 QY 1621 TGAAGGAGATGCATATGTAACGCTCATATGCGCCATAGTCTGATTTGGAAGAAAGTAG 1680
 Db 1680 TGAAGGAGATGCATATGTAACGCTCATATGCGCCATAGTCTGATTTGGAAGAAAGTAG 1739
 QY 1681 CTTTCTGATGAAGAAAGTTGAGCTCAAGCTTACTTAAAGAAAGGTTATCTTACTTACC 1740
 Db 1740 CTTTCTGATGAAGAAAGTTGAGCTCAAGCTTACTTAAAGAAAGGTTATCTTACTTACC 1799
 QY 1741 TCCATCTCCAGCAGCAGATGTTTAAAGCAAAATCCAACTGGAGATGTCAGAGCTATTTTA 1800
 Db 1800 TCCATCTCCAGCAGCAGATGTTTAAAGCAAAATCCAACTGGAGATGTCAGAGCTATTTTA 1859
 QY 1801 CAATCTGCTGAAAGGGGAAAAACGAATTTCCACTCGTTCGACTTCCATATATGTTTGAGCA 1860
 Db 1860 CAATCTGCTGAAAGGGGAAAAACGAATTTCCACTCGTTCGACTTCCATATATGTTTGAGCA 1919
 QY 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTCATAGGATCATTTACCAATAAT 1920
 Db 1920 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTCATAGGATCATTTACCAATAAT 1979
 QY 1921 TAAATTTGCTTTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAG 1980
 Db 1980 TAAATTTGCTTTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAG 2039
 QY 1981 TTTGTTTGGCAGCATTAAGTACTAGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2040
 Db 2040 TTTGTTTGGCAGCATTAAGTACTAGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2099
 QY 2041 TGGATGGGCAATGCCAGTGAAGTGTGTTAGGAGAAAGAACCAACAGTGAAGATCCAAA 2100
 Db 2100 TGGATGGGCAATGCCAGTGAAGTGTGTTAGGAGAAAGAACCAACAGTGAAGATCCAAA 2159
 QY 2101 TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGAAACCTGCTGAGCCAGAGTCC 2160
 Db 2160 TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGAAACCTGCTGAGCCAGAGTCC 2219
 QY 2161 TCAAGTAGAGACTGAAAGGTTAGAGCCCAACTCAAGAAAGCAGAGTTTGTCTTGGCAA 2220
 Db 2220 TCAAGTAGAGACTGAAAGGTTAGAGCCCAACTCAAGAAAGCAGAGTTTGTCTTGGCAA 2279
 QY 2221 AGTAAAGGATTTCTAGTCTGAAAGCCATGCAACGAAACTCTAGCTGTTTACGAAATAA 2280
 Db 2280 AGTAAAGGATTTCTAGTCTGAAAGCCATGCAACGAAACTCTAGCTGTTTACGAAATAA 2339
 QY 2281 TTTGACTCTTCAATTTATGATAACAAATAGTATCATGCGAGAGCAGAAAAATTACTTGC 2340

Db 2340 TTGACTCTTCAATATTGATGATACCAATAGTATCATGCGAAGAGCGAAAAATTAATCTTGC 2399

QY 2341 GTTGTAAAGGAAGTAATCTTCACTGTGTAAGTAGGAAAAATAAAC 2389

Db 2400 GTTGTAAAGGAAGTAATCTTCACTGTGTAAGTAGGAAAAATAAAC 2448

RESULT 7

AF291695 2541 bp DNA linear BCT 15-MAR-2001

LOCUS Streptococcus pneumoniae pneumococcal histidine triad A protein

DEFINITION (phtA) gene, complete cds.

ACCESSION AF291695

VERSION AF291695.1 GI:13345012

KEYWORDS Streptococcus pneumoniae

SOURCE Streptococcus pneumoniae

ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE 1 (bases 1 to 2541)

AUTHORS Wizenann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Maure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S., and Koenig, S.

TITLE Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection

JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)

MEDLINE 21116976

PUBMED 11179332

REFERENCE 2 (bases 1 to 2541)

AUTHORS Choi, G.H.

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences, Inc., 9410 Key West Ave., Rockville, MD 20850, USA

FEATURES

source location/Qualifiers

1..2541

/organism="Streptococcus pneumoniae"

/mol_type="genomic DNA"

/strain="N4"

/db_xref="taxon:1313"

91..2541

/gene="phtA"

91..2541

/gene="phtA"

/note="phtA"

/codon_start=1

/transl_table=11

/product="pneumococcal histidine triad A protein"

/protein_id="AAK19155.1"

/db_xref="GI:13345013"

/translation="MKINKYLVGSAALILSVCSYELGLYQARTVKENNRVSYIDGK QATQKNTPEVSRREGINAEQIVKITDQGYVSHGDHVVHYGKVPYDAIISEE LLMKDNYKLEEDIVNEVKGKGVIVKDGKYYVYLKDAHADNVRKEINRQKQHS QHREGPTPNDGVALARSQRTYTDGVIYFNASDIIEDTDAYIVPHGDHYHYPKN ELSASLAAEAFLSGRNLNLSRTYRQNSDTSNWPVSNFTNTNTSNNS TNSQASNDIISLQKLYKLPLSRHVSDELVPAPQITSTARGVAVPHGDHYH IFYQMSLEERLARIIPLYRNSHWVPSRPEOPSPPTPEPSGPQAPNLIKIDSN SSVLSOLVRKGVGEVFEKGISRYVFAKDLPSFTVKNLESKLKQESVSHULTAKE NVAPRQEFYDKRNYLLEHAKFLPNKGRNSDFQALKLERLNSINKEKLVDL LAPFTIPERIGKNSQLEYEDVEVRIQAADKYTSDGYIFDEHDIISDEGDVY TPMHGSWIGKDSLSDEKVAQAYTKEGILPPSPDADKANPTGDSAAAIYNRVK GKRIPLVRLPYMVEIVKNGNLIIPHKDHYNIKFAWDDHTYKAPNGYLEDLF ATIKYVEHPDERPHNDGWNASRHLGKDHSDPNKNFKADEPVBETPAEPEVP QVETKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLITLIQIWNDSIMAEKLL LALLKSNSSVSKEKIN"

gene

CDS

ORIGIN

Query Match 100.0%; Score 2388; DB 1; Length 2541;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTAGGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 209

61 TATAGATGGAACAACAGCGACGCAAAAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 120

210 TATAGATGGAACAACAGCGACGCAAAAAACCGAGAAATTTGACTCCTGATGAGGTAGCAA 269

121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGCGGTATGTCAC 180

270 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGCGGTATGTCAC 329

181 TTCACATGCGACCACTATCATTAATGTAAGGTTCCTTATGACGCTATCATCAG 240

330 TTCACATGCGACCACTATCATTAATGTAAGGTTCCTTATGACGCTATCATCAG 389

241 TGAAGATTACTCATGAAGATCCAACTATAGCTAAGATAAGAGATGAGATATCTTAATGA 300

390 TGAAGATTACTCATGAAGATCCAACTATAGCTAAGATAAGAGATGAGATATCTTAATGA 449

301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATATCTATGTTTACCTTAAGGATGC 360

450 GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATATCTATGTTTACCTTAAGGATGC 509

361 TGCCCAACGGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420

510 TGCCCAACGGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 569

421 TCAACATCGTGAAGGTGGAACTCCAGAAACGATGCTGTTGTCCTTGCATGAGGATGC 480

570 TCAACATCGTGAAGGTGGAACTCCAGAAACGATGCTGTTGTCCTTGCATGAGGATGC 629

481 AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540

630 AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 689

541 TGCTGATGCTTATATCGTCTCATCGAGATCATTAACCATTAATCTTCAAGAACTGTCAAA 600

690 TGCTGATGCTTATATCGTCTCATCGAGATCATTAACCATTAATCTTCAAGAACTGTCAAA 749

601 ATCAGCTAGCGAGTTCGCTGCTGACAGAGCTTCTTATCTGTCGAGGAAATCTGTCAAA 660

750 ATCAGCTAGCGAGTTCGCTGCTGACAGAGCTTCTTATCTGTCGAGGAAATCTGTCAAA 809

661 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAACTGTCGAGGATGC 720

810 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAACTGTCGAGGATGC 869

721 TGTAAGCAATCCAGGAACCTAACAATACTAACAAGCAACAAGCAACAAGCAACAAGCAACA 780

870 TGTAAGCAATCCAGGAACCTAACAATACTAACAAGCAACAAGCAACAAGCAACAAGCAACA 929

781 AGCAAGTCAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTTCAAACTGCTTTGAG 840

930 AGCAAGTCAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTTCAAACTGCTTTGAG 989

841 TCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCAACAATCAAGTCCGAAC 900

990 TCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCAACAATCAAGTCCGAAC 1049

901 AGCTAGAGTGTGAGTGGCAGCAGGATCATTAACCATTAATCTTCAAACTGCTTTGAG 960

1050 AGCTAGAGTGTGAGTGGCAGCAGGATCATTAACCATTAATCTTCAAACTGCTTTGAG 1109

961 GTCTGAATTTGGAAGAAGAAATCGCTCGTATTTATTCCTTCCCTTGGTATCGTTCAAACTG 1020

1110 GTCTGAATTTGGAAGAAGAAATCGCTCGTATTTATTCCTTCCCTTGGTATCGTTCAAACTG 1169

1021 GGTACAGATTTAAGGCCAGAAACAACAGTCCAAACCGACTCCGGAACTTAGTCCAGG 1080

1170 GGTACAGATTTAAGGCCAGAAACAACAGTCCAAACCGACTCCGGAACTTAGTCCAGG 1229

1081 CCGCGAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTCTGTTAGTCAAGTGGT 1140

Db 1230 CCGCAACCTGACCAAACTTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGTGGT 1289
QY 1141 AGGAAAGTTGGGGAAGATATGTAATTCGAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1200
Db 1290 ACGAAAGTTGGGGAAGATATGTAATTCGAAGAAAGGCGATCTCTCGTTATGTCTTTGC 1349
QY 1201 GAAAGATTACCATCTGAAACCTGTTAAATACTCTGAAAGCAAGTTATCAAAACAAGAGAG 1260
Db 1350 GAAAGATTACCATCTGAAACCTGTTAAATACTCTGAAAGCAAGTTATCAAAACAAGAGAG 1409
QY 1261 TGTCTTCAACACTTTAACTGCTTAAAGAAAGAAATGTGCTCTCGTGACCAAGAAATTTTA 1320
Db 1410 TGTCTTCAACACTTTAACTGCTTAAAGAAAGAAATGTGCTCTCGTGACCAAGAAATTTTA 1469
QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGGTGC 1380
Db 1470 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGGTGC 1529
QY 1381 TAAATTTCTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db 1530 TAAATTTCTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1589
QY 1441 TAAAGAAATTTGGTAGATGATTTATTTGCAATTCCTPAGCAACCAATTAACCATCCAGAGCG 1500
Db 1590 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTAACCATCCAGAGCG 1649
QY 1501 ACTTGGCAACCAAAATTTCTCAATTCAGTATGATGATGATGATGATGATGATGATGATGATG 1560
Db 1650 ACTTGGCAACCAAAATTTCTCAATTCAGTATGATGATGATGATGATGATGATGATGATGATG 1709
QY 1561 AGCTGATAGTATACAAACGCTCAGATGGTTACATTTTGTGATGAACATGATATATCAGTGA 1620
Db 1710 AGCTGATAGTATACAAACGCTCAGATGGTTACATTTTGTGATGAACATGATATATCAGTGA 1769
QY 1621 TGAAGAGATGATATGTAACGCTCATATGCGCCATAGTCCCATAGTCTGATGGAAGATAG 1680
Db 1770 TGAAGAGATGATATGTAACGCTCATATGCGCCATAGTCCCATAGTCTGATGGAAGATAG 1829
QY 1681 CTTTCTGATAGGAAAGTTGCGAGCTCAAGCCTATACAAAGAAAGGATCTCTACC 1740
Db 1830 CTTTCTGATAGGAAAGTTGCGAGCTCAAGCCTATACAAAGAAAGGATCTCTACC 1889
QY 1741 TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
Db 1890 TCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1949
QY 1801 CAATCGTGTGAAAGGGGAAAGAAAGAAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1860
Db 1950 CAATCGTGTGAAAGGGGAAAGAAAGAAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 2009
QY 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTATCTCTCATAGGATCATTAACATAATAT 1920
Db 2010 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTATCTCTCATAGGATCATTAACATAATAT 2069
QY 1921 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 2070 TAAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 2129
QY 1981 TTTGTTTGGACGATTAAGTACTACTAGAGAACACCTCTGACGACGTCACATTTCAATGA 2040
Db 2130 TTTGTTTGGACGATTAAGTACTACTAGAGAACACCTCTGACGACGTCACATTTCAATGA 2189
QY 2041 TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCAAGTGAAGATCCAA 2100
Db 2190 TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCAAGTGAAGATCCAA 2249
QY 2101 TAAAGAACTTCAAAGCGGATCAAGAGCCAGTAGAGAAACCTGCTGAGCCAGAAAGTCCC 2160
Db 2250 TAAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGAAACCTGCTGAGCCAGAAAGTCCC 2309
QY 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAAGAGCAGAAAGTTTGTTCGAA 2220
Db 2310 TCAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAAGAGCAGAAAGTTTGTTCGAA 2369

QY 2221 AGTAACGGAATTTAGTCTGAAAGCCAAATCCACAGAAACTCTAGCTGGTTAGGAATAA 2280
Db 2370 AGTAACGGAATTTAGTCTGAAAGCCAAATCCACAGAAACTCTAGCTGGTTAGGAATAA 2429
QY 2281 TTTGACTCTTCAAAATTTAGGATAACAAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2340
Db 2430 TTTGACTCTTCAAAATTTAGGATAACAAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2489
QY 2341 GTTGTAAAGGAAGTAATCTTCTCATCTCTGTAAGTAAGGAAAAATAAAT 2389
Db 2490 GTTGTAAAGGAAGTAATCTTCTCATCTCTGTAAGTAAGGAAAAATAAAT 2538

RESULT 8
AR218862
LOCUS AR218862 8195 bp DNA linear PAT 25-SEP-2002
DEFINITION Sequence 94 from patent US 6420135.
ACCESSION AR218862
VERSION AR218862.1 GI:23319796
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
Fannon, M.R., and Dougherty, B.A.
TITLE Streptococcus pneumoniae polynucleotides and sequences
JOURNAL Patent: US 6420135-A 94 16-JUL-2002;
FEATURES Location/Qualifiers
source 1..8195
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 2388; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60
Db 3053 TTTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 3112
QY 61 TATAGATGAAAAACAAGCGACGCAAAACACGAGAAATTTGACTCTCTGATGAGGTAGCAA 120
Db 3113 TATAGATGAAAAACAAGCGACGCAAAACACGAGAAATTTGACTCTCTGATGAGGTAGCAA 3172
QY 121 GCGTGAAGGAATCAATGTGTAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 180
Db 3173 GCGTGAAGGAATCAATGTGTAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 3232
QY 181 TTCACATGCGACCACTATCATTTACATGTTAAGGTTCCCTTATGACGCTATCATCAG 240
Db 3233 TTCACATGCGACCACTATCATTTACATGTTAAGGTTCCCTTATGACGCTATCATCAG 3292
QY 241 TGAAGAAATTTACTCATGAAAGATCCAAACTATTAAGCTAAAAGATGAGGATTTGTTAATGA 300
Db 3293 TGAAGAAATTTACTCATGAAGATCCAAACTATTAAGCTAAAAGATGAGGATTTGTTAATGA 3352
QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTTAAGGATGC 360
Db 3353 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTTAAGGATGC 3412
QY 361 TGCCCAACCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAAACAAGACGATAG 420
Db 3413 TGCCCAACCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAAACAAGACGATAG 3472
QY 421 TCAACATCGTGAAGGTGAACTCCAAAGAAACGATCGTGTGCTTGTGCTTGGCAGCTTCGA 480
Db 3473 TCAACATCGTGAAGGTGAACTCCAAAGAAACGATCGTGTGCTTGTGCTTGGCAGCTTCGA 3532
QY 481 AGGACGCTATACACAGATGATGTTATATCTTTTAAATGCTTCTGATATCATAGAGATAC 540

PF	30-OCT-1997 JP 1998520718	
PR	31-OCT-1996 US 60/029960	
PI	CHARLES A KUNSH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH,	
PI	MICHAEL FANNON, BRIAN A DOUGHERTY	
PC	C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,	
PC	C12N1/21,	
PC	C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC G06F15/40	
CC	Strandedness: Double;	
CC	Topology: Linear;	
PH	Key	
FT	source	
FT	Location/Qualifiers	
FT	1..8195	
FEATURES	Location/Qualifiers	
source	1..8195	
	/organism="unidentified"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32644"	
ORIGIN		
Query Match	100.0%; Score 2388; DB 6; Length 8195;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2388; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60	
Db	3053 TTCTTACGAGTTGGGACTCTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 3112	
QY	61 TATAGATGAAACAAAGCGACGCAAAAAACGGAGAATTTGACTCTCGATGAGGTAGCAA 120	
Db	3113 TATAGATGAAACAAAGCGACGCAAAAAACGGAGAATTTGACTCTCGATGAGGTAGCAA 3172	
QY	121 GGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATTAACAGCAACCAAGGCTATGTCAC 180	
Db	3173 GGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATTAACAGCAACCAAGGCTATGTCAC 3232	
QY	181 TTCACATGCGGACCACTATATTACAAATGTTAAGTAAAGATGAGGATTTGATGACGCTATCATCAG 240	
Db	3233 TTCACATGCGGACCACTATATTACAAATGTTAAGTAAAGATGAGGATTTGATGACGCTATCATCAG 3292	
QY	241 TGAAGAAATPACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATTTGTTAATGA 300	
Db	3293 TGAAGAAATPACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATTTGTTAATGA 3352	
QY	301 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC 360	
Db	3353 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC 3412	
QY	361 TGCCACGCGGATAGCTCGTACAAAGAGGGAATCAATCGCAAAAAACAGAGCATAG 420	
Db	3413 TGCCACGCGGATAGCTCGTACAAAGAGGGAATCAATCGCAAAAAACAGAGCATAG 3472	
QY	421 TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTGTTGCTTGGCAGCTTCGCA 480	
Db	3473 TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTGTTGCTTGGCAGCTTCGCA 3532	
QY	481 AGAAGCTATATCAAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGATAC 540	
Db	3533 AGAAGCTATATCAAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGATAC 3592	
QY	541 TGGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATTCATCTTAAAGATCAGTT 600	
Db	3593 TGGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATTCATCTTAAAGATCAGTT 3652	
QY	601 ATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCCTTATCTGTTGCGAGGAAATCTGTCAAA 660	
Db	3653 ATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCCTTATCTGTTGCGAGGAAATCTGTCAAA 3712	
QY	661 TTCAAGAACTTATCGCGCAAAAAATAGCGATAACATTTCAAGAAACAAACTGGGTACCTTC 720	
Db	3713 TTCAAGAACTTATCGCGCAAAAAATAGCGATAACATTTCAAGAAACAAACTGGGTACCTTC 3772	

QY	721 TGTAAAGCAATCCAGCACTACAAATACTAACAACAGCAACAGCAACACTTAACAGTCA 780	
Db	3773 TGTAAAGCAATCCAGCACTACAAATACTAACAACAGCAACAGCAACACTTAACAGTCA 3832	
QY	781 AGCAAGTCAAGTAAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAAACTGCTTTTGA 840	
Db	3833 AGCAAGTCAAGTAAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAAACTGCTTTTGA 3892	
QY	841 TCAACAGCATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAC 900	
Db	3893 TCAACAGCATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAC 3952	
QY	901 AGCTAGAGGTGTCAGTCCACACAGGAGATCATTTACCACTTCTATCTCTCAAT 960	
Db	3953 AGCTAGAGGTGTCAGTCCACACAGGAGATCATTTACCACTTCTATCTCTCAAT 4012	
QY	961 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTCGTTATCGTTTCAAAACCAT 1020	
Db	4013 GTCTGAATTTGGAAGAACGAATCGCTCGTATTATTTCCCTTCGTTATCGTTTCAAAACCAT 4072	
QY	1021 GGTACAGATTTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 1080	
Db	4073 GGTACAGATTTCAAGCCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG 4132	
QY	1081 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAAGTGT 1140	
Db	4133 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAAGTGT 4192	
QY	1141 ACGAAAGTTGGGAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTG 1200	
Db	4193 ACGAAAGTTGGGAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTG 4252	
QY	1201 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260	
Db	4253 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 4312	
QY	1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 1320	
Db	4313 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 4372	
QY	1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 1380	
Db	4373 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 4432	
QY	1381 TAAATCTGATTTCCAGGCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 1440	
Db	4433 TAAATCTGATTTCCAGGCTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAA 4492	
QY	1441 TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCAGAGCG 1500	
Db	4493 TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCAGAGCG 4552	
QY	1501 ACTTGGCAACCAAAATTTCTCAAAATTTGAGTATACTGAAGACGAAGTTTCGTTATGCTCAAT 1560	
Db	4553 ACTTGGCAACCAAAATTTCTCAAAATTTGAGTATACTGAAGACGAAGTTTCGTTATGCTCAAT 4612	
QY	1561 AGCTGATAGTATACAACTGAGATGTTTACATTTTGTATGAACATGATATATCAAGTGA 1620	
Db	4613 AGCTGATAGTATACAACTGAGATGTTTACATTTTGTATGAACATGATATATCAAGTGA 4672	
QY	1621 TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 1680	
Db	4673 TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 4732	
QY	1681 CCTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCTTATCTAAAGAAAAAGGATTCCTAAC 1740	
Db	4733 CCTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCTTATCTAAAGAAAAAGGATTCCTAAC 4792	
QY	1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGGAGAGCTATTATTA 1800	
Db	4793 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGGAGAGCTATTATTA 4852	
QY	1801 CAATCGTGTGAAAGGGGAAAAACGAATTCCTCGTTCGACTTCCATATATGTTGAGCA 1860	

Db	4853	CAATCGTGTGAAGGGGAAACGAATCCACTGTTCCGACTTCCATATATGTTGACCA	4912
Qy	1861	TACAGTTGAGGTTAAACCGTAATTTGATATTTCCTCATAGGATCATTTACCATATAT	1920
Db	4913	TACAGTTGAGGTTAAACCGTAATTTGATATTTCCTCATAGGATCATTTACCATATAT	4972
Qy	1921	TAAATTTGCTTGGTTTTCATGATACACATACAAAGCTCCAAATGCTATACCTTGGAGA	1980
Db	4973	TAAATTTGCTTGGTTTTCATGATACACATACAAAGCTCCAAATGCTATACCTTGGAGA	5032
Qy	1981	TTTCTTTGCGACGATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCAATGA	2040
Db	5033	TTTCTTTGCGACGATTAAGTACTACGTAGAACACCCCTGACGAAGCTCCACATTTCAATGA	5092
Qy	2041	TGGATGGGGCNATCCGAGTACGATGTTAGGCAAGAACACACAGTGAAGATCCAAA	2100
Db	5093	TGGATGGGGCNATCCGAGTACGATGTTAGGCAAGAACACACAGTGAAGATCCAAA	5152
Qy	2101	TAAGAACTTCAAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	5153	TAAGAACTTCAAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	5212
Qy	2161	TCAAGTAGAGACTGAAAAGTGAAGAGCCCAACTCAAAGAGAGCAAGTTTGTTCGCA	2220
Db	5213	TCAAGTAGAGACTGAAAAGTGAAGAGCCCAACTCAAAGAGAGCAAGTTTGTTCGCA	5272
Qy	2221	AGTAACGGATTCTAGTCTGAAGCCATGCAACGAACTCTAGCTGGTTACGAATAA	2280
Db	5273	AGTAACGGATTCTAGTCTGAAGCCATGCAACGAACTCTAGCTGGTTACGAATAA	5332
Qy	2281	TTTGACTCTTCAAAATTGGATACAAATAGTATCATGCGAAGCAGAAATAATTAAC	2340
Db	5333	TTTGACTCTTCAAAATTGGATACAAATAGTATCATGCGAAGCAGAAATAATTAAC	5392
Qy	2341	GTGTTTAAAGGAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC	2389
Db	5393	GTGTTTAAAGGAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC	5441
RESULT 10			
AE007418/c			
LOCUS	10256 bp	DNA	linear
DEFINITION	Streptococcus pneumoniae TIGR4	section 101 of 194 of the complete genome.	BCT 31-AUG-2001
ACCESSION	AE007418	AE005672	
VERSION	AE007418.1	GI:14972649	
KEYWORDS	Streptococcus pneumoniae TIGR4		
SOURCE	Streptococcus pneumoniae TIGR4		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
REFERENCE	1 (bases 1 to 10256)		
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.		
TITLE	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae		
JOURNAL	Science 293 (5529), 498-506 (2001)		
MEDLINE	21357209		
PUBMED	11463916		
REFERENCE	2 (bases 1 to 10256)		
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A.,		
source	1. 10256		
organism	"Streptococcus pneumoniae TIGR4"		
mol_type	"genomic DNA"		
strain	"TIGR4"		
db_xref	"taxon:170187"		
complement	(100..702)		
gene	"SP1170"		
CDS	complement(100..702)		
	gene="SP1170"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75279.1"		
	db_xref="GI:14972650"		
	/translation="MSEVDNEAVNYEFTSDTCLANSIYQSLPKPDKKNFSGDLIP		
	TKSPSLVKEGDYIGRRDSQVNLNRIYFNPNYLNRYKSLNMRNGCMGDFPHDF		
	DYLDHVAKYAYEQKNNIKYKYLKAILHQENALYFRFFSNDFDLEKNYKTIWQ		
	VSKETPSEMDENKXNISEKIIIFERSGKMLNDLKSNNK"		
	complement(760..1473)		
gene	gene="SP1171"		
CDS	complement(760..1473)		
	gene="SP1171"		
	/notes="identified by match to PFAM protein family HMM		
	PF00702"		
	/codon_start=1		
	/transl_table=11		
	/product="hydrolase, haloacid dehalogenase-like family"		
	/protein_id="AAK75280.1"		
	db_xref="GI:14972651"		
	/translation="MFYKFLFDLDTLLDFAADVALTQLLKEGVADTQAYKDY		
	VPNKALWKDLKTKISKVLNTRFSLRFAHFGQKDGSLAQRYQFYLAQOQSILS		
	GAHDLDSLIERDYNLYAATNGITATCTGLAQSLGAPYENOVFISLOLTKEDALF		
	YKIQGIQAGFSKREKTLMDGSLTADIQGNAGIDITWNPHEHLENTQAQTYEY		
	SYQDLDDLDKNIKLEKITF"		
	complement(1724..1843)		
gene	gene="SP1172"		
CDS	complement(1724..1843)		
	gene="SP1172"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75281.1"		
	db_xref="GI:14972652"		
	/translation="MLYAVFPYFNRSYIYFNCESIKTDCDGAILETFKN"		
	complement(1933..2307)		
gene	gene="SP1173"		
CDS	complement(1933..2307)		
	gene="SP1173"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved hypothetical protein"		
	/protein_id="AAK75282.1"		
	db_xref="GI:14972653"		
	/translation="MESIGLIVIVSHKHAIEGVVELISKVAKDVPIYVVRGTEGGIG		
	TSPEQVDRVVSSENPAITLAFDLSGAKNLMKMTDFSDKSIINRVPIVEGAYNAAA		
	LLQAGLSVICTQLAELEINK"		
	complement(2729..5188)		
gene	gene="SP1174"		
CDS	complement(2729..5188)		
	gene="SP1174"		
	/notes="similar to GP:4249624; identified by sequence		
	similarity; putative"		
	/codon_start=1		

Qy	961	GTCTGAATTGGAAGCAACGATCGCTCGTATTATTCCTTCCTGTTATCGTTCAAAACCATG	1020
Db	6777	GTCTGAATTGGAAGCAACGATCGCTCGTATTATTCCTTCCTGTTATCGTTCAAAACCATG	6718
Qy	1021	GGTACAGATTCAAGGCCAGAACCAACCAAGTCCACAACCGACTCCGGAACTAGTCCAGG	1080
Db	6717	GGTACAGATTCAAGGCCAGAACCAACCAAGTCCACAACCGACTCCGGAACTAGTCCAGG	6658
Qy	1081	CCGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTTCAGCTGGT	1140
Db	6657	CCGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTTCAGCTGGT	6598
Qy	1141	ACGAAAGTTGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGC	1200
Db	6597	ACGAAAGTTGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCTTTGC	6538
Qy	1201	GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAG	1260
Db	6537	GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAAGTTATCAAAACAAGAG	6478
Qy	1261	TGTTTCACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGATTTTA	1320
Db	6477	TGTTTCACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGATTTTA	6418
Qy	1321	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG	1380
Db	6417	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTTTGNAAATAAGGGTCG	6358
Qy	1381	TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTGTAATGATGATCGACTAA	1440
Db	6357	TAAATCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTGTAATGATGATCGACTAA	6298
Qy	1441	TAAAGAAAAATGGTAGATGATTTATTTGCCATTTCTTAGCAACCAATTACCCTCCAGAGCG	1500
Db	6297	TAAAGAAAAATGGTAGATGATTTATTTGCCATTTCTTAGCAACCAATTACCCTCCAGAGCG	6238
Qy	1501	ACTTGGCAACCAATCTCAATTCAGTATAGTGAAGACGAAGTTGCTATGCTCAAT	1560
Db	6237	ACTTGGCAACCAATCTCAATTCAGTATAGTGAAGACGAAGTTGCTATGCTCAAT	6178
Qy	1561	AGCTGATAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTGA	1620
Db	6177	AGCTGATAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTGA	6118
Qy	1621	TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACATGATTTGAAAAAGATAG	1680
Db	6117	TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACATGATTTGAAAAAGATAG	6058
Qy	1681	CCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTTATATAAGAAAAAGGTATCCTACC	1740
Db	6057	CCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTTATATAAGAAAAAGGTATCCTACC	5998
Qy	1741	TCCATCTCCAGACCGAGATGTTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA	1800
Db	5997	TCCATCTCCAGACCGAGATGTTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA	5938
Qy	1801	CAATCGTGTGAAGGGAAAAACGAATTCCTCGTTCGATTCCTATATATGTTTGAGCA	1860
Db	5937	CAATCGTGTGAAGGGAAAAACGAATTCCTCGTTCGATTCCTATATATGTTTGAGCA	5878
Qy	1861	TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATACCTAATAT	1920
Db	5877	TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATACCTAATAT	5818
Qy	1921	TAAATTTGCTTGGTTTGGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA	1980
Db	5817	TAAATTTGCTTGGTTTGGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA	5758
Qy	1981	TTTGTGTCGACGATTAAGTACTAGTAGAACACCTTGACGAACGTCACATCTTAATGA	2040
Db	5757	TTTGTGTCGACGATTAAGTACTAGTAGAACACCTTGACGAACGTCACATCTTAATGA	5698
Qy	2041	TGGATGGGGCAATGCCAGTGAGCATGTGTGTAGGCAAGAAACACCAAGTGAAGATCCAAA	2100
Db	5697	TGGATGGGGCAATGCCAGTGAGCATGTGTGTAGGCAAGAAACACCAAGTGAAGATCCAAA	5638
Qy	2101	TAAAGACTTTCAAAGCGGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	5637	TAAAGACTTTCAAAGCGGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCCC	5578
Qy	2161	TCAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAACGAGAAAGTTTCTTTCGAA	2220
Db	5577	TCAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAACGAGAAAGTTTCTTTCGAA	5518
Qy	2221	AGTAACCGATTTCTAGTCTGAAAGCAATGCAACGAAACTCTAGCTGTTTACGAAATAA	2280
Db	5517	AGTAACCGATTTCTAGTCTGAAAGCAATGCAACGAAACTCTAGCTGTTTACGAAATAA	5458
Qy	2281	TTTGACTCTTCAAAATTTAGTATACCAATAGTATATCATGCGCAAGCAGAAAAATTTCTGC	2340
Db	5457	TTTGACTCTTCAAAATTTAGTATACCAATAGTATATCATGCGCAAGCAGAAAAATTTCTGC	5398
Qy	2341	GTTGTTAAAGCAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC	2389
Db	5397	GTTGTTAAAGCAAGTAACTCTTCATCTGTAAGTAAAGGAAAAATAAAC	5349
RESULT 11			
AX571763/c			
LOCUS	AX571763	349980 bp	DNA linear PAT 29-NOV-2002
DEFINITION	Sequence 4982 from Patent WO02077021.		
ACCESSION	AX571763		
VERSION	AX571763.1	GI:26003955	
KEYWORDS	Streptococcus pneumoniae		
SOURCE	Streptococcus pneumoniae		
ORGANISM	Streptococcus pneumoniae		
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
Streptococcus.			
REFERENCE			
1			
Masignani, V., Tettelin, H. and Fraser, C.			
Streptococcus pneumoniae proteins and nucleic acids			
Patent: WO 02077021-A 4982 03-OCT-2002;			
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)			
FEATURES			
source			
1. 349980			
Location/Qualifiers			
/organism="Streptococcus pneumoniae"			
/mol type="unassigned DNA"			
/db xref="taxon:1313"			
/note="seq 4979 too long: 2.162.598 bases-replaced by			
following seq:-seq 4979: from 0.000.001 to 0.349.980-seq			
4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001			
to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq			
4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001			
to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq			
4986: from 2.100.001 to 2.162.598"			
ORIGIN			
Query Match 100.0%; Score 2388; DB 6; Length 349980;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTGTTCCTA	60
Db	106950	TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTGTTCCTA	106891
Qy	61	TATAGATGGAACCAACGACGCAACCAACGAGAAATTTGACTCCTCATGAGGTAGCAA	120
Db	106890	TATAGATGGAACCAACGACGCAACCAACGAGAAATTTGACTCCTCATGAGGTAGCAA	106831
Qy	121	CGGTGAAGGAATCAATCTGAGCAAAATCGTTCATCAAGATAACAGCAAGCTATGTCAC	180
Db	106830	CGGTGAAGGAATCAATCTGAGCAAAATCGTTCATCAAGATAACAGCAAGCTATGTCAC	106771
Qy	181	TTCAATGGCGACCACTATCATTTATTAACATGTAAGGTTCCCTTATGACGCTATCATCAG	240
Db	106770	TTCAATGGCGACCACTATCATTTATTAACATGTAAGGTTCCCTTATGACGCTATCATCAG	106711

[illegible]

[illegible]

[illegible]

Db	18630	TAGCTGATAAGTATACAAAGTACAGATGGTTTACATTTTGTATGAAACATGATATAAATCAGTG	18689
Qy	1620	ATGAAGCAGATGCATATGTAAACGCTCATATGGCCCATAGTCACTGGATTTGAAAAAGATA	1679
Db	18690	ATGAAGCAGATGCATATGTAAACGCTCATATGGCCCATAGTCACTGGATTTGAAAAAGATA	18749
Qy	1680	GCCTTTCTGATAGGAAAAAGTTTGCAGCTCAAGCCTATATCTAAAGAAAAAGGTATCCCTAC	1739
Db	18750	GCCTTTCTGATAGGAAAAAGTTTGCAGCTCAAGCCTATATCTAAAGAAAAAGGTATCCCTAC	18809
Qy	1740	CTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATT	1799
Db	18810	CTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATT	18869
Qy	1800	ACAATCTGTGAAAGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATATGTTGAGC	1859
Db	18870	ACAATCTGTGAAAGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGC	18929
Qy	1860	ATACAGTTGAGGTTTAAAAACGGTAATTTGATTTATCTCATAGGATCAATTAACATAATA	1919
Db	18930	ATACAGTTGAGGTTTAAAAACGGTAATTTGATTTATCTCATAGGATCAATTAACATAATA	18989
Qy	1920	TAAATTTGCTTGGTTTGATGATCACACATACAAAAGCTCCAAATGCTATATCTTGGAG	1979
Db	18990	TAAATTTGCTTGGTTTGATGATCACACATACAAAAGCTCCAAATGCTATATCTTGGAG	19049
Qy	1980	ATTTGTTTGGCAGCAATTAAGTACTAGTGAACACACCTGACGAACGCTCCACATCTCTAATG	2039
Db	19050	ATTTGTTTGGCAGCAATTAAGTACTAGTGAACACACCTGACGAACGCTCCACATCTCTAATG	19109
Qy	2040	ATGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAA	2099
Db	19110	ATGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAA	19169
Qy	2100	ATAGAACTTCAAGCCGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCC	2159
Db	19170	ATAGAACTTCAAGCCGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCC	19229
Qy	2160	CTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTGGCA	2219
Db	19230	CTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTGGCA	19289
Qy	2220	AAGTAACGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA	2279
Db	19290	AAGTAACGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATA	19349
Qy	2280	ATTTGACTCTCAAAATTTATGATAACATAGTATCATGCGAAGACGAGAAAAATTTACTTG	2339
Db	19350	ATTTGACTCTCAAAATTTATGATAACATAGTATCATGCGAAGACGAGAAAAATTTACTTG	19409
Qy	2340	CGTTGTTAAAGGAAGTAATCCTTCATCTCTAAGTAAGGAAAAATAAAC	2389
Db	19410	CGTTGTTAAAGGAAGTAATCCTTCATCTCTAAGTAAGGAAAAATAAAC	19459
RESULT 14			
AF340221			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

MEDLINE 21246685
PUBMED 11349048
REFERENCE 2 (bases 1 to 2535)
AUTHORS Zhang, Y., Masi, A., Barniak, V., Mountzouros, K., Hostetter, M. and Green, B.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES
Location/Qualifiers
1..2535
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
1..2535
/gene="pnpA"
1..2535
/gene="pnpA"
/notes="histidine motif-containing protein"
/codon_start=1
/transl_table=11
/product="PnpA"
/protein_id="AAK26629.1"
/db_xref="GI:113447094"
/translation="MKINKKYLVSAAALILSVCSYELGLYQARTVKNNRVSYIDGK QATQKENTLPDEVRKREGINAEQIVIKITDQGVTSHGDIHYVYNGKVPYDAIISSE LMKDPYOLKDEDIISSEIKGVYKVDGKYVYLKDAHADNVRKBEINRQKQSHS OHREGGTPRNDGVALARSQGVTTDDGVYENASDIIEDTDGAIIVPHGDHYVYIPKN ELASLAARAKAPLSRGNLSRSTRYRONSNTSRNWPVSNGPTNTNTSNNNS TNSQASQNDISLLKOLPLUSQRHVESDGLIFDPAQITSRKTANGVAVPHGDHYHF IYPSQSLPEEKLIARIPIYLRNSHWVPSRPEQSPQSPSPSPAPNPQAPPS NPIDEKLKVEARFVNGDGVFEENGVPYIPAKLSAETANGIDSLKQESLSHKL AKKDLPPSSDRFVNGYDILLARIHODLNDKGRQVDPEALNDLRLKQVSDSKVL VDDILAFAPIRHPEBLRPNQITVYDDEIOVAKLGVYTTEDGVIIPDRIDTSDEG DAVTEPHMTHSHWIKDLSSEARAAQVAKELGTPSPDHOQSENTEAKGAERAY NRUKAAKVPDRMPNLYOYVEVKGSLIIIPHYDHNIKFEWFEDEGLYEAQKVSLL EDLLATVEHPNRSHDNGFNASDHVQRNKGQADNTQTEKPNKPEKQTEKPE EETPRKEPQSEKPEKPEPEESPEEPQTEKPEKVEKLEAREADLLGK IQNPIIKSNAKETLITLGNLLFGTDQNNITMAEAEKLLALLKESK"

Query Match 60.0%; Score 1434.4; DB 1; Length 2535;
Best Local Similarity 79.0%; Pred. No. 5.6e-296;
Matches 1724; Conservative 0; Mismatches 447; Indels 12; Gaps 1;

QY 1 TTCTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60
DB 60 TTCTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 119
QY 61 TATAGATGAAACAAAGCGACGCAAAACCGGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
DB 120 TATAGATGAAACAAAGCGACGCAAAACCGGAGAAATTTGACTCCTGATGAGGTTAGCAA 179
QY 121 GCGTGAAGCAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
DB 180 GCGTGAAGCAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 239
QY 181 TTCACATGCGGACCACTATCATTTATCAATGTTAAGGTTCTTATGAGGCTATCATCAG 240
DB 240 TTCACATGCGGACCACTATCATTTATCAATGTTAAGGTTCTTATGAGGCTATCATCAG 299
QY 241 TGAAGAAATTAATCATGAAGATCCAAACCTAAGCTAAAGATGAGGATATTGTTAATCA 300
DB 300 TGAAGAAATTAATCATGAAGATCCAAACCTAAGCTAAAGATGAGGATATTGTTAATCA 359
QY 301 GGTCAAGGCTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 360
DB 360 ATCAAGGCTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 419
QY 361 TGCCACGCGGATAACGTCCTGATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 420
DB 420 AGCTCATGCGGATAAATGTCCTGATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG 479
QY 421 TCACATCATGTAAGGTGGAACTCCAGAAACGATGCTGTGCTGTTCCTGGCAGGTTCCGCA 480

DB 480 TCACATCATGTAAGGTGGGACTCCAGAAACGATGGTGTGCTGTTCCTGGCAGGTTCCACA 539
QY 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 540 GGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 599
QY 541 TGTGTATGCTTATATCGTTCCTCATGAGATCATTAACCATTAATCATTCCTTAAGATGAGTT 600
DB 600 TGTGTATGCTTATATCGTTCCTCATGAGGACCATTAACCATTAATCATTCCTTAAGATGAGTT 659
QY 601 ATCAGCTAGCGAGTGGCTGTGCAAGACCTTCTTATCTGTGTCGAGGAAATCTGTCAA 660
DB 660 ATCAGCTAGCGAGTGGCTGTGCAAGACCTTCTTATCTGTGTCGAGGAAATCTGTCAA 719
QY 661 TTCAAGAACCTATCGCGGACAAATAAGCGATACACTTCAAGAACAACTGGGTACTTTC 720
DB 720 TTCAAGAACCTATCGCGGACAAATAAGCGATACACTTCAAGAACAACTGGGTACTTTC 779
QY 721 TGTAAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA 780
DB 780 TGTAAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA 839
QY 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTTGAACAGCTCTACAAACTGCTTTGAG 840
DB 840 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTTGAACAGCTCTACAAACTGCTTTGAG 899
QY 841 TCACGACATGATAGTAATCTGTATGAGGCTTCTTTGATCCAGCAACAAATCAAGTCCGAA 900
DB 900 TCACGACATGATAGTAATCTGTATGAGGCTTCTTTGATCCAGCAACAAATCAAGTCCGAA 959
QY 901 AGCTAGAGTGTGGCAGTCCACAGGAGATCATTAACCACTTCACTCCCTTACTCTCAAA 960
DB 960 CGCCAAATGCTGTGTTACCGCAGGACCACTATCACTTTATCTTATTTCACTACT 1019
QY 961 GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTTTCGTTATCGTTCAAAACATTG 1020
DB 1020 GTCACTTTTGAAGAAATAATGGCTCGTATTTATTCCTTTCGTTATCGTTCAAAACATTG 1079
QY 1021 GGTACCAAGATTCAGCCAGAACCAAGTCCACAGCACTCCGGAACCTAGTCCAGG 1080
DB 1080 GGTACCAAGATTCAGCCAGAACCAAGTCCACAGCACTCCGGAACCTAGTCCAGG 1139
QY 1081 CCGCAACCTCGACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT 1128
DB 1140 TCGCAACCTCGACCAAAATCTTCAACCACTCCAGCAATCCAAATGATGAGAAATGGT 1199
QY 1129 TAGTCAGCTGTGACGAAAGTTGGGAGAGGATATGTTATTCGAGAAAGAGGATCTCTCG 1188
DB 1200 CAAAGAAAGCTGTTCGAAAGTAGGCGATGGTTATGTTTGGAGAGATGGAGTTCTCTCG 1259
QY 1189 TTATGCTTTTGGAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATC 1248
DB 1260 TTATATCCAGCAAGGATCTTTTCAGCAGAAACAGCAGCAGGATTTGATAGAACTGGC 1319
QY 1249 AAAACAAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATTTGTGCTCTCGTCA 1308
DB 1320 CAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTGACCTCCCATCTAGTGA 1379
QY 1309 CCAAGAAATTTATGATAAGCATATACTGTTTAACTAGGCTCATAAAGCCTTGTGTTGN 1368
DB 1380 TAGAGAAATTTACAATAGGCTTATGACTTACTAGCAAGAAATTCATCAAGATTTACTTGA 1439
QY 1369 AAATAAGGCTCGTAAATCTGATTTTCCAGCCTTTAGACAAATTTATAGACGCTTGAATCA 1428
DB 1440 TAATAAGGCTCAAGAGTTGATTTTGAAGCTTTGGATATACCTGTTGGAAACGACTCAAGGA 1499
QY 1429 TGAATCGAATAAAGAAATAATGGTAGATGATTTATTTGGCAATTTCTAGACCAATTAAC 1488
DB 1500 TGTCTCAAGTGATAAAGTCAAGTCTAGTGGATGATATTTCTGCTTCTTAGCTCCGATTCCG 1559
QY 1489 CCATCCAGAGCGACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGTTCCG 1548

[illegible]

1484 TAATAAAGGTCGACAAAGTTGATTTGGGCTTTGGATPAACCTGTTGGACACTCAAGGA 1543
 1429 TGAATCGACTAATTAAGAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACCAATPAC 1488
 1544 TGTCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTTCTGCTTCTTAGCTCGGATTCG 1603
 1489 CCATCCAGAGCGACTTGGCAACCAATTTCTCAAAATGAGTATCTGAAGACGAAGTTTCG 1548
 1604 TCATCCGAGAGTTTAGGAAACCAATTCGCAATTTACCTACACTGATGATGAGTTCA 1663
 1549 TATTGCTCAATTAGCTGATAGTATACAACTGATGTTTACATTTTGGATGAACTGA 1608
 1664 AGTAGCCAAGTTGCGAGCAAGTACACACAGAGAGCGTTATATCTTTGATCTCTCGTGA 1723
 1609 TATTAATCAGTATGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGAT 1668
 1724 TATAACAGTATGAGGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTGGAT 1783
 1669 TGGAAAGATAGCCCTTTCTGATAAGGAAAGTTGACAGTCCAGCTCAAGCCTATCTAAAGAAA 1728
 1784 TAAAGAGATAGTTTGTCTGAGCTGAGAGCGCGACGCCAGGCTTATGCTAAAGAGAA 1843
 1729 AGTATCCTACCTCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGC 1788
 1844 AGGTTTGACCCCTCCTTCGACAGACCATCAGATTTCAGGAAATACTGAGGCAAAAGGAGC 1903
 1789 AGCAGCTATTTACAACTGTTGAAGGGGAAAGAAAGAAATCCAACTGGTTCGACTTCCATA 1848
 1904 AGAAGCTATCTACAAACCGCTGAAAGCAGCTTAAAGGTTGCGCACTTGTATGCTGCTTA 1963
 1849 TATGGTTGAGCATACAGTTCGAGTTTAAAGCGTAAATTTGATTTATCTCTCATAGGATCA 1908
 1964 CAATCTTCAATATCTAGAGTCAAAACCGGTAGTTTATCATCTCATATGACCA 2023
 1909 TTACCATATATTAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTA 1968
 2024 TTACCATATATCAAAATTTGAGTGGTTGACGAGGCTTTATGAGGCACTTAAGGGGTA 2083
 1969 TACCTTGGAGATTTGTTGCGAGATTAAGTACTAGTAGAGACACCTGACGAGCTCC 2028
 2084 TACTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACTCC 2143
 2029 ACATTTCAATGATGATGGGCAATGCGAGTACATGTGTTAGGCAAGAAAGACCAAG 2088
 2144 GCATTGAGATTAATGTTTGTAACTGAGTACGACCATGTTCAAGAAACAAATAATGCTCA 2203
 2089 TGAAGTCCAAATGAAGAACTTCAAGCGGATGAAG - 2125
 2204 AGCTGATACCAATCAAAACGGAAACCAAGCGAGGAAACCTCAGACAGAAACCTGA 2263
 2126 ----- 2125
 2264 GGAAGAAACCCCTCGAGAGAGAAACCAAAAGCGAGAAACCGAGTCTCCAAACCAAC 2323
 2126 -----CCAGTAGAGGAAACACCTGCTGAGCAGAGAGTCCCTCAAGTAGAGACTGAAA 2178
 2324 AGAGAACCAAGAGAAATCACAGAGGATCAGAGAACTCAGTTCAGACTGAAA 2383
 2179 AGTAGAAGCCCACTCAAGAGAGAGAGTTTGGTTGCGAAAGTAAACGGAATCTAGTCT 2238
 2384 GGTGGAAGAAACCTGAGAGAGGCTGAAGATTTACTTTGGAAAAATCCAGGATCCAAATTA 2443
 2239 GAAAGCCAAATGCAACAACTCTAGTGGTTTACGAAATTAATTTGACTCTTCAATAT 2298
 2444 CAGTCAATGCGCAAGAGACTCTCAGAGATTTAAAAAATAATTTACTATTTGGCACCCA 2503
 2299 GGATAACAATAGTATCATGCGAGAGAGAGAAATTAATTTGCGTTGTTAAAAAGAGATA 2358
 2504 GGACAACTACTATTTATGCGAGAGCTGAAACACTATTGGCTTTTAAAGAGAGATA 2563

AX343073
 LOCUS AX343073 2647 bp DNA linear PAT 12-JAN-2002
 DEFINITION Sequence 4 from Patent WO0198334.
 ACCESSION AX343073
 VERSION AX343073.1 GI:18152271
 KEYWORDS
 SOURCE unidentified
 ORGANISM Streptococcus antigens
 unclassified.
 unclassified.
 1
 REFERENCE
 AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
 TITLE Streptococcus antigens
 JOURNAL Patent: WO 0198334-A 4 27-DEC-2001;
 SHIRE BIOCHEM INC. (CA)
 FEATURES
 source
 1. 2647
 Location/Qualifiers
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /note="BVH-11"
 ORIGIN
 Query Match 57.5%; Score 1374.2; DB 6; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 4.1e-283;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGCTTAAGGAAATTAATCGTGTTCCTA 60
 Db 104 TCGTTATGAATAGTGTTCATCAAGCTCAACTGTAAAGAAATTAATCGTGTTCCTA 163
 Qy 61 TATAGATGGAAGAAACCAACGACGCAAAACCGAGAAATTTGACTCCTCATGAGGTTAGCAA 120
 Db 164 TATAGATGGAAGAAACCAACGACGCAAAACCGAGAAATTTGACTCCTCATGAGGTTAGCAA 223
 Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAACGAGCTATGTCA 180
 Db 224 GCGTGAAGGAATCAACGCGCAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTAC 283
 Qy 181 TTCATATGGCGACCATATCATTTATTAATGTTAGGTTCCCTTATGACGCTATCATCAG 240
 Db 284 CTCTCATGGAGACCATTTATCTACTATAATGCGAGGTCCTTTATGATGCCATCATCAG 343
 Qy 241 TGAAGAAATTAATCATGAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA 300
 Db 344 TGAAGACTCCTCATGAAGATCCGAATTTACGTTGAAGGATTCAGACATTTGCAATGA 403
 Qy 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
 Db 404 AATCAAGGGTGGTTATGTCATTAAAGGTAAACGGTAAATCTATGTTTACCTTAAGGATGC 463
 Qy 361 TGCCCAAGCGGATTAACGTCGTTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420
 Db 464 AGCTCATGCGGATTAATGTCGTTACAAAGAGGAAATCAATTCGGCAAAACAAAGACATAG 523
 Qy 421 TCAACATCTGTGAAGGTGGAATCTCAAGAAACGATGTTGTTGCTTGGCTTGGCAAGTTGCGA 480
 Db 524 TCAGATCTGTGAAGGAGGAGCTTCAGCAACGATGGTGGGTAGCCCTTTTCACGTTTACA 583
 Qy 481 AGGACGCTACTACTACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 Db 584 GGGACGCTACACACAGATGATGTTTATCTTCAATGATCTGATATCATCGAGATAC 643
 Qy 541 TGTGATGCTTATATGTTTCTCATGAGATCATTTACCATTTACATTCCTTAAGATGAGTT 600
 Db 644 GGGCGATGCTATATGTTTCTCATGAGATCATTTACCATTTACATTCCTTAAGATGAGTT 703
 Qy 601 ATCAGCTAGCGACTTGCCTGCTGACAGCCTTCTCTATCTGCTGAGGAAATCTGTCAA 660
 Db 704 ATCAGCTAGCGAGTTGGCTGCTGACAGGCTTCTCTATCTGCTGCGGAAATCTGTCAA 763
 Qy 661 TTCAAGAACTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTCCTTC 720
 Db 764 TTTAAGAACTATCGCCGACAAAATAGCGATACACTTCAAGAAACAACTGGGTCCTTC 823


```

|||||
1938 CCATTACCAATCAATCAATTTGAGTGTGTGACGAAGCCCTTATGAGGCACCTAAGGG 1997
|||||
1966 CTATACCTTGAAGATTTTGTGGACGATTAAGTACTAGTAGAACCCCTGACGAACG 2025
|||||
1998 GTATACCTTGGAGATCTTTTGGGACGTCAAGTACTATGTGCAACATCAAAACGAACG 2057
|||||
2026 TCCACATTTCTAATGATGATGGGCAATGCCAGTGAGCATGTGTAGGCAACGAAGACCA 2085
|||||
2058 TCCGCATTTACGATAATGTTTGTGTAACGCTAGCGACCATGTTCAAGAAACAAAATGG 2117
|||||
2086 CAGTGAAGATCCAAATAAGAACTTCAAAAGCGGATCA----- 2121
|||||
2118 TCAAGCTATACCAATCAACCGAANAACCAAGCGAGGAGAACCTCAGACAGAAACCC 2177
|||||
2122 ----- 2121
|||||
2178 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGCGAGAACACAGAGTCTTCCAAAACC 2237
|||||
2122 ---AGAGCCAGTAGAGGAACACACCTGCTGAGCGAGAGTCCCTCAAGTAGAGACTGAATA 2178
|||||
2238 AACAGAGGAACCAAGAAGATACCAAGAGATCAGAGAACCTCAGTTCGAGACTGAATA 2297
|||||
2179 AGTAGAAGCCCACTCAAGAGAGAGAGATTTTGTTCGGAAGTAACGGATTTCTAGTCT 2238
|||||
2298 GGTGAAGAAAACTCAGAGAGGCTGAAGATTTACTTGGMAAAATCCAGGATCCAAATTAT 2357
|||||
2239 GAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAATTTGACTCTTCAAAATTAT 2298
|||||
2358 CAAATCCCAATGCAAGAGACTCTCACAGGATTAATAATAATTTACTATTTGGCACCCA 2417
|||||
2299 GGATACCAATAGTATCATGCGAGAGAGAGCAAAAAATTTACTTGGTGTGTTAAAGGAAGTAA 2358
|||||
2418 GGACAAACAATATATTATGGCAGAGAGCTGAAAAACTATTGGCTTTTAAAGGAGATGA 2477
|||||

RESULT 20
BD264091
LOCUS 2481 bp DNA linear PAT 17-JUL-2003
DEFINITION Streptococcus pneumoniae proteins and nucleic acid molecules.
ACCESSION BD264091
VERSION BD264091.1 GI:33073859
KEYWORDS JP 2002531055-A/27.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 2481)
Gilbert,C.F.G. and Hansbro,P.M.
Streptococcus pneumoniae proteins and nucleic acid molecules
Patent: JP 2002531055-A 27 24-SEP-2002;
MICROBIAL TECHNIQS LTD
OS Streptococcus pneumoniae
PN JP 2002531055-A/27
PD 24-SEP-2002
PF 27-JUL-1999 JP 2000562519
PR 27-JUL-1998 GB 9816337.1 19-MAR-1999 US 60/125164 PI
CHRISTOPHE FRANCOIS GUY GILBERT, PHILIP MICHAEL HANSBRO PC
C12N15/09, C12N15/09, A61K39/09, A61K45/00, A61K48/00, A61P31/04, PC
C07K14/315,
PC C07K16/12, C12Q1/14, C12Q1/68, G01N33/53, G01N33/566, G01N33/569//
PC C12P21/08,
PC G01N33/15, G01N33/50, (C12N15/09, C12R1:46), (C12Q1/14, C12R1:46),
PC C12N15/00,
PC C12N15/00, (C12N15/00, C12R1:46)
CC Streptococcus pneumoniae proteins and nucleic acid molecules
FH key Location/Qualifiers
FT source 1. .2481
FT /organism='Streptococcus pneumoniae'.
FT Location/Qualifiers
1. .2481
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
FEATURES
source
|||||
958 AATGCTGAATTGGAAGAACGAATCGCTGATATTATCCCTTCCTGTTATCGTTTCAACCA 1017
|||||

```

```

/db_xref="taxon:1313"
ORIGIN
Query Match 42.0%; Score 1003.8; DB 6; Length 2481;
Best Local Similarity 66.0%; Pred. No. 5.8e-204; Indels 144; Gaps 6;
Matches 1623; Conservative 0; Mismatches 693;
QY 1 TTCTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCTGTTTTC 57
DB 60 TTCCTATGAGCTTGGTGGTCAACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGC 119
QY 58 CTATATAGATGGAAAAACAAGCGACGCAAAAAACGAGAAATTGACTCTCTGATGAGTTAG 117
DB 120 TTATATAGATGGTGTATCAGGCTGGTCAAAAGGCGAAAAAAGTTCACACAGATGAAGTCAG 179
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATTAACAGATTAACAGACCAAGGCTATGT 177
DB 180 TAAGAGGAGGGGATCAACGCGCAACAAATCGTCATCAAGATTAACAGATTAACAGGTTATGT 239
QY 178 CACTTCATCGCGACCACTATCAATTAATCAATGGTAAAGTTCTTATGACGCTATCAT 237
DB 240 GACCTCTCATGGAGCACTTATCACTATAATGGCAAGGTCCCTTATGATGCCATCAT 299
QY 238 CAGTGAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAA 297
DB 300 CAGTGAAGAGCTCCCTCATGAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTTGTCAA 359
QY 298 TGAGGTCAAGGTTGATATGTTATCAAGCTAGATGGAATAATACTATGTTTACCTTAAAGGA 357
DB 360 TGAATCAAGGTTGTTATGATCATCAAGGTAGACGGAATAATACTATGTTTACCTTAAAGGA 419
QY 358 TGCTGCCACCGGATAACGTCGTAACAAAGAGGAAATCAATCGCAAAAAACAAGACA 417
DB 420 TGCACCTCATCGGATAATATTTCGCAAAAAAGAGAGATTAACCGTCAGAAGCAGGAAC 479
QY 418 TAGTCAACATCGTGAAGGTGGAACCTCCAGAAACCATGCTGCTGTTGCTTGGCAGCTTC 477
DB 480 CAGTCATAATCACGGGTCAGGAGCT-----AACGATCATGCAAGTAGCTGCGCCAGAGC 533
QY 478 GCAAGGAGCTATATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGA 537
DB 534 CCAAGGAGCTATATCAACGGATGATGGGTATATCTTCAATGATCTGATATCATTTAGAGA 593
QY 538 TACTGGTGTATGCTTATATCGTTCCCTCATGAGATCATTAACCATTAATTCCTTAAGATGA 597
DB 594 CACGGGTGATGCTTATATCGTTCCCTCAAGCGGACCATTAACCATTAATTCCTTAAGATGA 653
QY 598 GTTATCAGCTAGCAGTTGGCTGCTGCAAGAGCTTCTCTATCTGCTGCGAGGAATCTGTC 657
DB 654 GTTATCAGCTAGCAGTTAGCTGCTGCAAGAGCTTATTGGAATGG----- 698
QY 658 AAATTCAGAAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
DB 699 -GAAGCAGGATCTCGTCTTCTTCAAGTTCTAGTTAATGCAATCCAGCTCAACCAA 757
QY 718 TTCTGTGAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACAGCAACACTAACAG 777
DB 758 GATTTGTGAGAGAACCAATCT-----GACTGTCACTCCAATTA 797
QY 778 TCAAGCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAACTGCCTTT 837
DB 798 TCATCAAAATCAAGGGGAAAACATTTCAAGCCCTTTTACGTGAATGTTGTTAAACCTTT 857
QY 838 GAGTCAACACACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCACAAAATCAAGTCG 897
DB 858 ATCAGAACCCATGTGGATCTGATGGCTTTATTTTCGACCCAGCGCAAAATCAAGTCG 917
QY 898 AACAGCTAGAGTGTGAGTGCCACACGAGATCATTAACCATTTATCCCTTACTCTCA 957
DB 918 AACCGCCAGAGGTGATGCTGCTCCCTCATGTAACCATTAACCTTTATCCCTTATGAACA 977
QY 958 AATGCTGAATTGGAAGAACGAATCGCTGATATTATCCCTTCCTGTTATCGTTTCAACCA 1017
|||||

```

Db 978 AATGCTCTGAATTGGAAGAAACGAATTGCTGTAATTAATTCCTCCCTTCGTTATCGTTCAAAACCA 1037
QY 1018 TTGGGTACCAAGATTCAAGGCGCAGAACCAAGTCCACCAACCGACTCGGACCTAGTCC 1077
Db 1038 TTGGGTACCAAGATTCAAGACAGAACCAAGTCCACCAACCGACTCGGACCTAGTCC 1097
QY 1078 AGGCCCGCAACCTGCACCAATCTTAAATATAGACTCAA-----ATTCTCTCTT 1125
Db 1098 AAGTCCGCAACCTGCACCAATCTCAACAGCTCCAGCAATCCAAATTGATGAGAAATT 1157
QY 1126 GGTATAGTCACTGGTACGAAAGTTGGGGAAGATATGTATTCGAAAGAAAGGCACTTC 1185
Db 1158 GGTCAAAAGAGCTGTTTCAAAAGTAGCGATGGTATGTCTTTGAGGAGAAATGGAGTTTC 1217
QY 1186 TCGTTATGCTTTTGGCGAAGATTACCATCTGAAACTGTTAAATCTTGAAGCAAGTT 1245
Db 1218 TCGTTATCCAGCCAGGATCTTTTCAGCAGAAACAGCAGCGCAATGATAGCAAACT 1277
QY 1246 ATCAAAACAAAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCCTCG 1305
Db 1278 GGCCAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG 1337
QY 1306 TGACCAAGAAATTTATGATAAGCATATAATCTGTTAACTAGAGCTCATAAAGCCTTGT 1365
Db 1338 TGATCGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTACT 1397
QY 1366 TGAATAAAGGGTCGTAATTCGATTTCCAGGCTTAGCAAAATTTATGAAGCTTGAA 1425
Db 1398 TGATAATAAGGTCGACAAAGTTGATTTGAGGCTTTGGATAACCTGTGGAACGACTCAA 1457
QY 1426 TGATGAATCGACTAATAAGAAAAATTTGGTAGATGATTTATGGCATTCCTAGCACCAT 1485
Db 1458 GGATGTCCCAAGTGATAAGTCAAGTTAGTGGATGATATTCCTGCTTCTTAGCTCGAT 1517
QY 1486 TACCCATCCAGCGGACTGGGAAACCAATCTCAATTTAGTATGATCTAGCAAGCAAGT 1545
Db 1518 TCGTATCCAGAAAGCTTTAGGAAACCAATGGCCAAATTTACCTACACTGATGATGAT 1577
QY 1546 TCGTATTTGCTCAATTAGCTGATAAGTATACAACTGATGATGATGATGATGATGATGAT 1605
Db 1578 TCAAGTAGCAAGTTGGCAGGAAAGTACAAAGAGAGCGTTATATCTTTGATCCTCG 1637
QY 1606 TGATATAATCAGTATGAAAGGAGATGATATGATGATGATGATGATGATGATGATGATGAT 1665
Db 1638 TGATATAACCAAGTATGAGGAGGATGCTATGATGATGATGATGATGATGATGATGATGAT 1697
QY 1666 GATTGAAAGATAGCTTCTGATAGGAAAGTTGAGCTCAAGCTCATAGCTTACTAAGA 1725
Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAGA 1757
QY 1726 AAAAGGTATCTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAG 1785
Db 1758 GAAAGTTTGACCCCTCTCGACAGACCATCAGGATTCAGGAAATFACGAGGCAAAAG 1817
QY 1786 TGACAGCAGCTATTACAATCGTGTGAAAGGGGAAAAACGAATTCCTACTGTTTCCGCTTC 1845
Db 1818 AGCAGAAAGCTATCTACAACCGGTGAAAGCAGCTAAGAGGTCCTCACTTATGATGATGCC 1877
QY 1846 ATATATGGTTGAGCATACAGTTGAGGTTAAACCGGTAAATTTGATTTATTCCTCATAGA 1905
Db 1878 TTACATCTTCAATATCTGTAAGTCAAAACCGGTAGTTTAAATCATACCTCATTTATGA 1937
QY 1906 TCATTACCAATATTAATTTTCTGTTGTTGATGATACACATACAAAGCTCCCAATCG 1965
Db 1938 CCATTACCATACATCAAAATTTGAGTTGTTGAGGAGGCTTTATGAGGCACTTAAGG 1997
QY 1966 CTATACCTTGGAGATTTGTTTGGCAGATTAAGTACTAGTAAACACCTGACCAAGC 2025
Db 1998 GTATACTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGCGAATCTCAACCAAGC 2057
QY 2026 TCCACATTTCAATGATGGGATGGGCAATGCCAGTGAAGATGTTTGAAGCAAGAGACCA 2085
Db 2058 TCCGCAATTCAGATATGTTGTTTGGTAACGTAGCGACCATGTTTCAAGAAACCAAAATGG 2117

QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAAAGCGGATGA----- 2121
Db 2118 TCAAGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
QY 2122 ----- 2121
Db 2178 TGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAAAGCGAGAAACGAGTCTCCAAAACC 2237
QY 2122 ---AGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2238 AACAGAGAAACCCAGAGAATCACAGAGGAATCAGAGAACCTCAGCTCAGACTGAAAA 2297
QY 2179 AGTAGAAGCCCACTCAAGAGAGAGAGTTTTCGTTGCGAAAGTAAAGGATCTTAGTCT 2238
Db 2298 GGTGGAAGAAAACTGAGAGAGGCTGAAGATTACTTTGGAATAATCCAGGATCCAAATTAT 2357
QY 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAATTTGACTCTTCAATATAT 2298
Db 2358 CAAGTCCATGCCAAAGAGACTCTCAGGATTTAAAAATAATTTACTATTGGCACCCA 2417
QY 2299 GGATAACAATAGTATCATGCGAGAGCAGAAAAATTTACTTGGCTTTGTTAAAAAGGAAGTAA 2358
Db 2418 GGACAACTACTATTATGCGCAGAGCTGAANAATTTATGCTTTATTAAGGAGAGTAA 2477

RESULT 21
AX569137 AX569137 2457 bp DNA linear PAT 29-NOV-2002
LOCUS Sequence 2345 from Patent WO02077021.
DEFINITION
ACCESSION AX569137
VERSION AX569137.1 GI:26002635
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1
AUTHORS Masignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2345 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES
source
1..2457
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"

ORIGIN

Query Match 41.6%; Score 993; DB 6; Length 2457;
Best Local Similarity 65.8%; Pred. No. 1.2e-201;
Matches 1611; Conservative 0; Mismatches 696; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGGACTGTATCAGCTAGAACGGTTTAAGGAAAA---TAATCGTCTTTC 57
Db 60 TTCTTATGAGCTTGACGTTTACAGCTGCTCAGGATAAGAGAGTCTTAATCGAGTTGC 119
QY 58 CTATATAGATGAAAAACAAAGCGACCGCAAAAAACGGAGAAATTTGACTCTCTGATGAGTTAG 117
Db 120 TTATATAGATGTTGATCAGCTGGTGGTCAAAAGCGCAGAAAACTTGACACAGATGAAGTCAG 179
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATCTCATCAAGATAACAGACCAAGGCTATGT 177
Db 180 TAAGAGGGAGGGGATCAACGCGCAAAATTTGTTATCAAGATTACGGATCAAGGTTATGT 239
QY 178 CACTTCAATGCGGACCACTATCATTTTACAACTGTAAGGTTTCCTTATGAGCTATCAT 237
Db 240 GACCTCTCATGAGACCATTAATCTATTAATGGAAGGTTCCCTTAGTGCCATCAT 299
QY 238 CAGTGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATTTGTTAA 297
Db 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTGACATTTGTCAA 359

QY 298 TGAGGCTCAAGGCTGGATATCTTATCAAGGTAGATGGAAATACTATGTTTACCTTAAAGGA 357
 Db 360 TGAATCAAGGCTGGTATGTATTAAGGTAAACGGTAATACTATGTTTACCTTAAAGGA 419
 QY 358 TGCTGCCACCGCGATTAACGTCCTTACAAAAGAGAGAAATCAATCGACAAAACAAAGAGCA 417
 Db 420 TGCAGCTCATCGCGATTAATTCGACAAAGAGAGATTAACGTCAGAGCAGGAACG 479
 QY 418 TAGTCAACATCGTGAAGGTGAATCTCCAAAGAAACGATGGTCTGTGCTGGCAGCTTC 477
 Db 480 CAGTCATAATCAT-----AACTCAAGAGCAGATAATGCTGTGTCGAGCCAGAGC 530
 QY 478 GCAGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGGA 537
 Db 531 CCAAGGAGCTTATCAACGGATGATGGGTATATCTTCAATGCACTGATATCATTTGAGGA 590
 QY 538 TACTGGTGATCTTATATCGTTTCTCATGGAGATCATTAACCATTAATTCCTAAGAATGA 597
 Db 591 CACGGTGATCTTATATCGTTCTCCACGGCGACCATTAACCATTAATTCCTAAGAATGA 650
 QY 598 GTTATCAGTACGAGTTGGCTGTGCGAGAGCCTTCTATCTGTCGAGGAATCTGTC 657
 Db 651 GTTATCAGTACGAGTTAGCTGTGCGAGAGCCTTATGGAATGG----- 695
 QY 658 AAATTCAGAACCTTATCCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
 Db 696 -GAAGCAGGATCTCGTCTTCTCAAGTTCTAGTTTAAAGTAAAGTCAAGCTCAACCAA 754
 QY 718 TTCTGTAAAGCAATCCAGGAATCTACAAATACTAACACAGAACACACAGCAACACTAACAG 777
 Db 755 GATTGTCAAGAACCAACATCT-----GACTGTCACTCCCACTTA 794
 QY 778 TCAAGCAAGTCAAGTAATGATGATGATGATCTCTTTGAAACAGCTCTACAACTGCCWTT 837
 Db 795 TCATCAAAATCAAGGGGAAAACATTTCAAGCCCTTTTACGTGAATTTGATGCTAAACCCCTT 854
 QY 838 GAGTCAACGACATGTAGAACTGTATGCTGTCTTGTATGATCAGACACAAATCACAAAGTCG 897
 Db 855 ATCAGAACCCATGTGGAACTGTATGAGCTTATTTTGAACCCAGCGCAATCACAAAGTCG 914
 QY 898 AACAGCTAGAGGTGTGAGTGCCACACGGAGATCAATACCACTTCACTTCTTACTCTCA 957
 Db 915 AACCCCAAGAGGTGTAGCTGTCCCTCATGTAAACCAATTAACCACTTATCCCTTATGAACA 974
 QY 958 AATGTCTGAATTTGGAAGACGATCGCTGTATTAATCCCTTCTGTTATCGTTTCAACCA 1017
 Db 975 AATGTCTGAATTTGGAAGACGAAATGCTGTATTAATCCCTTCTGTTATCGTTTCAACCA 1034
 QY 1018 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 1077
 Db 1035 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 1094
 QY 1078 AGGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTCAGCT 1137
 Db 1095 AAGTCCGCAACCACTCCAAAGCAATCCAATTTGATGAGAA-----ATTGCTCAAAAGAGC 1148
 QY 1138 GGTACGAAAGTTGGGAGGATATGATTTCCAGAAAGGCACTCTCTGTTATGTCCT 1197
 Db 1149 TGTTCGAAAGTAGGCGATGGTTATGCTTTGAGGAGATGAGTTTCTGTTATATCCC 1208
 QY 1198 TCGCAAGAAATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGA 1257
 Db 1209 AGCCCAAGGATCTTTCAGCAGAACACAGCAGCAGCATTTGATCAAACTGGCCAGCAGGA 1268
 QY 1258 GAGTGTTCACACATTTTAACTGTCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAAT 1317
 Db 1269 AAGTTTATCTATAAGCTAGGAACCTAAGAAACCTGACCTCCCATCTAGTGTGCGAAT 1328
 QY 1318 TTATGATAAGCATATAATCTGTTAACTGAGGCTCAATAAGCCTTTTGNAAATAAGGG 1377
 Db 1329 TTACATAAGGCTTATGCTTACTAGCAAGAAATTCACCAAGATTTACTTGAATAAAGG 1388
 QY 1378 TCGTAATCTGATTTCCAGCCTTAGACAAATTTATAGAACGCTTGAATGATGATCGAC 1437

RESULT 22

Db 1389 TCGCAAGATTGATTTTGAGGCTTTGCGATAACCTGTTGGAAACGACTCAAGGATGTTCTCAAG 1448
 QY 1438 TAATAAGAAATAATTTGATGATGATTTTATTTGGCAATCTTCCAGCACAATTTACCCATCAGA 1497
 Db 1449 TGATAAAGTCAAGTTAGTGGAAAGATAATTTGCTTCTTTAGCTCCGATTTGCTATCAGA 1508
 QY 1498 GCGACTTTGGCAAAACCAAAATCTCAAAATGAGTATCTGAAGACGAAGTTCTGATTTGCTCA 1557
 Db 1509 ACGTTTAGGAAACCAAAATGCGCAAAATTAACCTACATGATGATGATGATGATGATGATGAT 1568
 QY 1558 ATTAGCTGATAAGTATACAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1617
 Db 1569 GTTGGCAGGCAAGTACACACAGAAAGACGGTTATATCTTTGATCTCTGATGATGATGATGATGAT 1628
 QY 1618 TGATGAAGGAGATGATATGTAAGCCTCATATGCGCATAGTCACTGCAATGTTGAAAGA 1677
 Db 1629 TGATGAGGGGATGCTTATGTAACCTCAATGACCCATGACCCATGACCCATGACCCATGACCCAT 1688
 QY 1678 TAGCCTTTCTGATAAGGAAAGTTGAGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 1737
 Db 1689 TAGTTTGTCTGAAGCTGAGAGCGCGAGCCAGCTTATGCTAAAGAGAAAGGTTTTCAC 1748
 QY 1738 ACCTCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGAGATAGTGGCAGAGCTAT 1797
 Db 1749 CCCTCTTCGACAGACCATCAAGATTCAGGAAATTAAGGCAAAAGAGAGAGAGCTAT 1808
 QY 1798 TTACAATCGTGTGAAGGGGAAACGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1857
 Db 1809 CTACAACCGCTGAAAGCAGCTTAAGAGTGCCACTTGTATGATGCTGCTGCTGCTGCTGCTGCTGCT 1868
 QY 1858 GCATACAGCTTGAAGGTTTAAAGACGGTAAATTTGATTTTCTCTAATAGGATCAATACCATAA 1917
 Db 1869 ATATCTGTGAAGTCAAAACCGTAGTTTAAATCATACCTCATATGATGATGATGATGATGATGAT 1928
 QY 1918 TATTAATTTGCTGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
 Db 1929 CATCAAAATTTGAGTGGTTTGACGAAGGCTTTATGAGGACCTTAAGGGGATGATGATGATGAT 1988
 QY 1978 AGATTGTTTTCGACGATTAAAGTACTAGTGAACACCTGACGAACTGACGAACTGACGAACTGAC 2037
 Db 1989 GGAATCTTTGGGAGTGTCAAGTACTGTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCA 2048
 QY 2038 TGATGATGCGGCAATGCGAGTGAAGTGTGTTAGGCAAGAAAGACCAAGTGAAGATCC 2097
 Db 2049 TAAATGTTTGTGAACGCTAGCGACCATGTTCAAGAAACCAAAATGCTCAAGCTGATAC 2108
 QY 2098 AATTAAGAACTTCAAGCGGATGA----- 2121
 Db 2109 CAATCAAAACGAAACCAACGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAAC 2168
 QY 2122 -----AGAGCGAGT 2130
 Db 2169 CCCTCGAAGAGAAACCGAAACCGAGAAACCGAGTCTCCAAACCAACAGAGAAAC 2228
 QY 2131 AGAGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAGTAGAGCCCA 2190
 Db 2229 AGAAGAAATCACCAGAGGAATCAGAAAGACCTCAGGTGAGACTGAAAGGTTGAAAGAAA 2288
 QY 2191 ACTCAAGAGAGCAGAGTTTTCCTTCCGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2250
 Db 2289 ACTGAGAGGCTGAAGATTTTACTTTGAAAGAAATCCAGGATCCAAATATCAAGTCCAAATG 2348
 QY 2251 AACGAAACTCTAGCTGTTTACGAAATAAATTTGACTCTTCAAAATTTATGGATAACAAATAG 2310
 Db 2349 CAAAGAGACTCTCAGAGGATTAAGAAATTAATTTACTTATTTGCGCCAGCAGCAACAAATAC 2408
 QY 2311 TATCATGCGAGAGCAGAGAAATTTACTTCCGTTTGTAAAGGAAGTAA 2358
 Db 2409 TATTATGCGAGAGCTGAAAACTATTGCTTTTATTAAGAGGAGTAA 2456

QY	838	GAGTCAACGACATGTAGAATCTGATCGCCTTGTCTTTGATCGAGCACAAATCACAAAGTCG	897
Db	855	ATCAGAACGCCATGTGGAATCTGATGCCCTATTTTCGACCACGCCAAATCACAAAGTCG	914
QY	898	AACAGCTAGAGGTGTTCAGTGCACACGAGATCAATACCACTTCATCCCTTACTCTCA	957
Db	915	AA CGCCAGAGGTGTAGCTGTCCCTCATGTGAACCAATTACCACTTTATCCCTTATGAACA	974
QY	958	AATGCTCGAATTGGAAGAAAGAAATCGTTCGTATATTTCCCTTCGTTATCGTTCAAAACA	1017
Db	975	AATGCTCGAATTGGAAGAAAGAAATCGTTCGTATATTTCCCTTCGTTATCGTTCAAAACA	1034
QY	1018	TTGGGTACCAGATTCAGAGGCCAGAAACAACCAAGTCACAACCGACTCCGGAACCTAGTCC	1077
Db	1035	TTGGGTACCAGATTCAGACACAGAGAACCAAGTCACAACCGACTCCGGAACCTAGTCC	1094
QY	1078	AGGCCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCT	1137
Db	1095	AAGTCCGCAACAGCTCCAAGCAATCCAATGTGATGGAA-----ATTGGTCAAGAAGC	1148
QY	1138	GGTACGAAAGTTGGGAAGGATATGATTCGAAGAAAGGCGCATCTCTGTTATCTCTT	1197
Db	1149	TGTTTCGAAAGTAGCGGATGTTAATGTTTCTTTGAGGAGATGGAGTTTCTCGTTATATCC	1208
QY	1198	TGCGAAAGATTTACCATCTGAAACTGTAAATCTTTGAAAGCAAGTTATCAAAACAAG	1257
Db	1209	AGCCAAGATCTTTCAGCAGAAACAGCAGCAGCGCATTTGATGACAACTGCGCCAAAGCAGGA	1268
QY	1258	GAGTGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCTCGTGACCAAGAAAT	1317
Db	1269	AAGTTTATCTCATAGCTTAGGAACCTAAGAAACTGACCTCCCATCTAGTATCGAGAAT	1328
QY	1318	TTATGATAAGCATATATCTGTAACTGAGGCTCATAAAGCCCTGTTGNAATTAAGG	1377
Db	1329	TTACAATAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACTTGATATAAGG	1388
QY	1378	TCGTAATTCGATTTCCAGCCTTAGACAAATTAATAGAACGTTGATGATGAATCGAC	1437
Db	1389	TCGACAAGTTGATTTGAGGCTTTGGATAACCTGTGGAACGACTCAAGGATGCTCAAG	1448
QY	1438	TAATAAGAAAAATTTGGTGTAGATGATTTATTTGGCATCTTAGCACTTACCATCCAGA	1497
Db	1449	TGATAAGCTCAAGTTAGTGGGAAGATATTTCTTGCTTTCTTAGTCCGATTCGTATCCAGA	1508
QY	1498	CGCACTTGGCAAAACCAATCTCAAAATGAGTATACCTGAAGACGAAGTTCGTATGCTCA	1557
Db	1509	ACGTTTAGAAAAACCAATGCGCAATTTACCTACACTGATGAGATTCAGGTAGCCAA	1568
QY	1558	ATTAGCTGATAGTATACAACTGATGAGTGTATCAATTTTGATGAAACATCATATATTCAG	1617
Db	1569	GTTGCGAGCGAAGTACACAGCAGAGACGGTTATATCTTTGATCCTCGTATATAACCG	1628
QY	1618	TGATCAAGGAGATGCATATGTAAACGCTCATATGGGCGATAGTCACCTGGATTTGGAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTGGATTTAAAAAGA	1688
QY	1678	TAGCCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATCTATAAGAAAAAGGTATCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGCGCGACCGCGCTTATGCTTGAAGAGAAAGGTTTGAC	1748
QY	1738	ACCTCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTCGACGAGCTAT	1797
Db	1749	CCCTCTTCGACAGACCATCAGGATTCAGGAATATCTGAGCAAAAGGACGAGAGCTAT	1808
QY	1798	TTACAATCGTGTGAAGGGGAAAAACGAATTCCTCGTTCGATTCCTATATGTTGA	1857
Db	1809	CTACAACGMGTGTAAGCAGCTTAAGAAGTGCCACTTGATCGTATGCTCTACATCTTCA	1868
QY	1858	GCATACAGTTGAGTTTAAAAACGGTAAATTTGATTTATCTCATAGGATCATTTACCATAA	1917
Db	1869	ATATACCTGTAGAGTCAAAAAACGGTAGTTTATCATACCTCATTTAGCACTTACCATAA	1928

QY	1918	TATTAAATTTGCTTTGGTTTGATGATCAACATACAAAGCTCCAAATGGCTTATACCTTGG	197
Db	1929	CATCAAAATTTGAGTGGTTTTCACGAAGCCCTTTATAGAGCACCTTAAGGGGTATACTCTTTGA	1988
QY	1978	AGATTTCGTTTGGCAGCAATTAAGTACTACGTAGAACACCCCTGAGGAACGTCACCATTTCTAA	2037
Db	1989	GGATCTTTTGGCGACTCTGAGTACTATGTGCAACATCCAAACGACGTCGGCATTCAGA	2048
QY	2038	TGATGATGGGGCAATCCAGTCAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCC	2097
Db	2049	TAATGGTTTGGTAAACGCTAGCGACCATGTTCAAGAAGAAACAAAATGGTCAAGCTGATAC	2108
QY	2098	AAATGAAGACTTCAAAGCGGATGA	2121
Db	2109	CAATCAAAACGGAAAAACCAAGCGGAGAGAAACCTCAGACAGAAAAACCTGAGGAGAAAC	2168
QY	2122	-----AGAGCCAGT	2130
Db	2169	CCCTCGAAGAGAGAAACCGCAACGCGAGAGAAACAGAGTCTCCAAAACCAACAGAGGAAC	2228
QY	2131	AGAGGAAACACCTCTGAGCGAGAAAGTCCCTCAAGTAGAGACTGAAAAGGTAGAAGCCCA	2190
Db	2229	AGAAAGATCACAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAGGTTTGAAGAAA	2288
QY	2191	ACTCAAAGAAGCAGAAGTTTGTCTTCGGAAGTAACCGATTTCTAGTCTGAAAAGCCAATGC	2250
Db	2289	ACTGAGAGAGCTGAGAGATTACTTTGGAAAAATCCAGATCCCAATTATCAAGTCCAATGC	2348
QY	2251	AACAGAACTCTAGCTGGTTTACGAATAATTTGACTCTTCCTCAAAATTTAGGATAACAATAG	2310
Db	2349	CAAAAGAGACTCTCAGAGATTAAAAATAATTTACTATTTGGCACCAGGACCAACATAC	2408
QY	2311	TATCATGCGAAGCAGAAAAATTTACTTGGCTGTGTTAAAGGAAGTAA	2358
Db	2409	TATTATGCGAAGCTGAAAACTATTGGCTTTTAAAGGAGAGTAA	2456
RESULT 23			
BD268049			
LOCUS	BD268049	2531 bp	DNA linear PAT 17-JUL-2003
DEFINITION	Streptococcus pneumoniae protein and immunogenic fragments for vaccines.		
ACCESSION	BD268049		
VERSION	BD268049.1	GI:33077817	
KEYWORDS	JP 2002532561-A/7.		
SOURCE	Streptococcus pneumoniae		
ORGANISM	Streptococcus pneumoniae		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 2531)		
AUTHORS	Johnson,L.S., Koenig,S. and Adamou,J.E.		
TITLE	Streptococcus pneumoniae protein and immunogenic fragments for		
JOURNAL	Patent: JP 2002532561-A 7 02-OCT-2002;		
	MEDIMUNE INC		
COMMENT	OS Streptococcus pneumoniae		
	PN JP 2002532561-A/7		
	PD 02-OCT-2002		
	PR 21-DEC-1999 JP 2000589215		
	PF 21-DEC-1998 US 60/113048		
	PI LESLIE S JOHNSON,SCOTT KOENIG,JOHN E ADAMOU		
	PC A61K39/09,A61K38/00,A61P31/10,A61K37/02		
	CC Streptococcus pneumoniae protein and immunogenic fragments for		
CC	vaccines		
key	Location/Qualifiers		
PH	1..2531		
FT	/organism='Streptococcus pneumoniae'.		
FT	Location/Qualifiers		
	1..2531		
	/organism='Streptococcus pneumoniae'		
	/mol_type='genomic DNA'		
	/db_xref='taxon:1313'		
FEATURES			
source			
ORIGIN			

Query Match		41.5%; Score 991; DB 6; Length 2531;
Best Local Similarity		65.7%; Pred. No. 3.2e-201;
Matches 1609, Conservative		1; Mismatches 697; Indels 141; Gaps 6;
Qy	1	TTCTTAGAGTGTGGAGTGTATCAAGCTAGACGGTTAAGGAAAA---TAATCGTGTTC 57
Db	60	TTCTATAGCTTGGACGTTTACCAAGCTGTGAGGATTAAGAGAGTCTAATCGAGTTGC 119
Qy	58	CTATATAGATGGAACACAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAG 117
Db	120	TTATATAGATGAGTATGATCAGGCTGTGCAAAAGGAGGAGAACTTGACACCATGAAGTCAG 179
Qy	118	CAAGCGTGAAGGAATCAATCTCAGCAAAATCGTCATCAAGATAACAGACCAAGCGCTATGT 177
Db	180	TNAGAGGAGGGATCAACGCCGACAAATGTTATCAAGATTACGATCAAGGTTATGT 239
Qy	178	CACCTCACATGGCGACCATATCATTATTAATAGTTAAGGTTCCCTTATGACGCTATCAT 237
Db	240	GACCTCTCATGGAGACCATATCATTTACTATATAATGGCAAGTTCCCTTATGATGCCATCAT 299
Qy	238	CAGTGAAGATTTACTCATGAAGATCCAACTATAGCTTAAGATAAGAGTGAAGATTTGTTAA 297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAAGTTTGAAGGATTCAGACATTTGTCAA 359
Qy	298	TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTTAAGGA 357
Db	360	TGAATCAAGGGTGGTTATGTCATTAAAGGTAAAGGTAAATACTATGTTTACCTTTAAGGA 419
Qy	358	TGCTGCCACGCGGATTAACGTCGTTCAAAAGAGGAAATCAATCGACAAAAACAAGACA 417
Db	420	TGCRGCTCATGCGATAATATTCGGACAAAGAGAGATTTAAAGCTCAGAGCAGGAACG 479
Qy	418	TAGTCAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGCTGCTTGGCTTGGCAAGTTC 477
Db	480	CAGTCAATATCAT-----AACTCAAGACGAGATAATGCTGTTGCTGCAGCCAGAGC 530
Qy	478	GCAAGGACGCTATPACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATACAGGA 537
Db	531	CCAAGGACGTTATACACGGATGATGGGTATATCTTCAATGCACTGATATCATTTGAGGA 590
Qy	538	TACTGTGATGCTTATATCGTTTCTCATGGAGATCATTAACATTAATCTTAAGAAATGA 597
Db	591	CACGGGTGATGCTTATATCGTTTCTCATGGACCATTAACATTAATCTTAAGAAATGA 650
Qy	598	GTATCAGCTAGCGTTCGCTGCTCAGAAAGCTTCTTATCTGTCGAGGAAATCTGTC 657
Db	651	GTATCAGCTAGCGTTCGCTGCTCAGAAAGCTTATGGAATGG----- 695
Qy	658	AAATTCAGAACTTATCGCCGACAAATAGCGATAACACTTCAAGAAACAACTGGGTACC 717
Db	696	-GAAGCAGGATCTCGTCTTCTTCAAGTTCCTAGTTTATTAATGCAATCCAGCTCAACCAA 754
Qy	718	TTCTGTAGCAATCCAGGAATCAAAATACCAAGCAACAACAGCAACACTAACAG 777
Db	755	GATTTGTGAGAGAACCAATCT-----GACTGTCACTCCAATTA 794
Qy	778	TCAAGCAAGTCAAGATTAATGATGATGATCTCTTGAACAGCTCTCAAACTGCCCTTT 837
Db	795	TCATCAAAATCAAGGGGAAAAATTTCAAGGCTTTTACGTGAATTTGTTGCTTAAACCCCTT 854
Qy	838	GAGTCAACGACATGTAGATCTGATGGCTTCTTGTGATCCAGCACAAATCACAAAGTCG 897
Db	855	ATCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAAGTCG 914
Qy	898	AACAGTAGAGTGTTCAGTGCACACAGGATCATTAACACTTATCCCTTACTCTCA 957
Db	915	AACCCGACAGAGTGTAGTGTCCCTCATGTTAACATTAACACTTATCCCTTATGAACA 974
Qy	958	AATGCTCAATTTGGAAGAACGATCGCTGTTATTTATCCCTTCTGTTATCGTTTCAAAACA 1017
Db	975	AATGCTCAATTTGGAAGAACGAAATGCTCGTTATTTATCCCTTCTGTTATCGTTTCAAAACA 1034

Qy	1018	TTGGTACCAAGATTCAAGGCCAGAAACAACAGTCCACAACCGACTCCGGAACCTAGTCC 1077
Db	1035	TTGGTACCAAGATTCAAGGCCAGAAACAAGTCCACAACCGACTCCAGAACCTAGTCC 1094
Qy	1078	AGGCCCAACCTGCGACCAATCTTAAATAGACTCAATCTTCTTCTTGGTGTAGTCACT 1137
Db	1095	AGTCCGCAACCAAGCTTCAAGCAATCCAATGATGGGAA-----ATTGTCAAAGAACG 1148
Qy	1138	GGTACGAAAGTTGGGAGAGGATATGTTATTCGAGAAAGAGGCACTCTCTGTTATGCTT 1197
Db	1149	TGTTGAAAGATGAGCGATGTTATGTTCTTTGAGGAGATGGAGTTCTCGTTATATCCC 1208
Qy	1198	TGCGAAGATTTACCATCTGAAAAGTGTAAATCTTTGAAAGCAAGTTATCAAAACAAGA 1257
Db	1209	AGCCAAGGATCTTTACGACGAAAACAGCAGCAGGCAATGATAGCAAACTGCCCAAGCAGGA 1268
Qy	1258	GAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGATTT 1317
Db	1269	AGTGTATCTCATAGCTAGGAACTAAGAAACTGACCTCCCATCTAGTATCGAAGATTT 1328
Qy	1318	TTATGATAAAGCATATAATCTGTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGG 1377
Db	1329	TTACAATAAGGCTTATGACTTACTAGCAAGATTTACCAAGATTTACTTGTATTAAGG 1388
Qy	1378	TGTAATTTCTGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGAC 1437
Db	1389	TCGACAGTGTGATTTTGGAGCTTTGGATAACCTGTTTGGAAACGACTCAAGGATGCTCAAG 1448
Qy	1438	TTATAAGAAATAATTCGTAGATGATTTATTGGCATTTCTAGCACCATAATACCCATCCAGA 1497
Db	1449	TGATAAGTCAAGTGTAGTGAAGATATCTTGCCTTTAGCTCCGATTCGTCATCCAGA 1508
Qy	1498	GGGACTTGGCAACCAATTTCTCAATGAGTATCTGAAGACGAAAGTTCGTATTGCTCA 1557
Db	1509	ACGTTTAGGAAACCAATTCGCAAAATTTACCTACACTGATGATGAGATTTCAAGTAGCCAA 1568
Qy	1558	ATTAGCTGATAGTATAGCAACGTCAGATGTTTACATTTTGTGATGAACATGATATAATCAG 1617
Db	1569	GTGGCAGCAAGTACACAGCAGAAAGCGTTTATCTTTGATCCTCGTATATAACCAAG 1628
Qy	1618	TGATGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGATGTTGGAAGA 1677
Db	1629	TGATGAGGGGATGCTATGTAATCCCATATGATGATGATGATGATGATGATGATGATGATGAT 1688
Qy	1678	TAGCCTTTCTGATAAGGAAAAAGTTGCAAGCTTATCTTAAGAAAAAGGATTCCT 1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCTRAAGAGAAAGGTTTAC 1748
Qy	1738	ACCTCCATCTCCAGAGCGAGATGTTAAAGCAATTCCACTGAGATGATGAGAGCTAT 1797
Db	1749	CCCTCCTTCGACAGACCATTCAGGATTTCAAGAAATATCTGAGGCAAGGAGAGAGCTAT 1808
Qy	1798	TTACAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCCGACTTCCATATATGTTGA 1857
Db	1809	CTACAACCGMTGGAAGCAGCTAAGAGGTGCCATCTGATGATGCTGCTTCAATCTTCA 1868
Qy	1858	GATACAGTGTAGGTTTAAACCGTAAATTTGATTTCTCTCAAGAGGATCATTAACATAA 1917
Db	1869	ATATATCTGTAGAAGTCAAAAACCGTAGTTTAAATCATACCTCATTTATGACCAATTAACATA 1928
Qy	1918	TATTAATTTCTGTTTGTGATGATCACAATCAAGCTCAAAATGGCTATACCTTGA 1977
Db	1929	CATCAAAATTTGAGTGTGAGAGGCTTTTATGAGGCACTTAAAGGCTTACTCTTGA 1988
Qy	1978	AGATTGTTTTCGCGAGTAAAGTACTAGTGAACACCTGACCAACCTGCCACATTTCTAA 2037
Db	1989	GGATCTTTTGGGACTGTCAAGTACTATGTCGAACATCCAAACGACCTCCGATTCAGA 2048
Qy	2038	TGATGATGGGCAATGCCAGTGAAGATGTTTGAAGCAAGAAACACACAGTGAAGATCC 2097
Db	2049	TAATGTTTTCGTAACGCTAGCGCAATGTTTCAAGAAACAAATGCTCAAGCTGATAC 2108
Qy	2098	AAATAGAACTTCAAGCGGATGA----- 2121


```

|||||
2109 CAATCAACGGAACCAACGAGGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAC 2168
2122 -----AGAGCCAGT 2130
2169 CCCTCGAAGAGAGAAACCGCAAGCGAGAAACACAGAGTCTCCAAAACCAACAGAGGAACC 2228
2131 AGAGGAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTCAAAAAGTAGAGGCCA 2190
2229 AGAAGATCAACGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAAGGTGAGAAAA 2288
2191 ACTCAAGAGACGAACTTTGCTTGGCAAGTAACCGATTCTAGTCTGAAAGCCAAATGC 2250
2289 ACTGAGAGAGCTGAAGATTACTTTGGAAAAATCCAGATCCAAATTAAGTCCAAATGC 2348
2251 AACAGAAACTCTAGCTGGTTTACGAATAATTTGACTCTTCAMATTAAGTAAACAATAG 2310
2349 CAAAGAGACTCTCAGAGATTAAAAAATAATTTACTATTATTGGCACCCAGGACAAATAC 2408
2311 TATCATGCGACAGACGAAAAATTTACTTGGCTTGTAAAGGAAGTAA 2358
2409 TATTATGCGACAGCTGAAAAACTATTGGCTTTTAAAGGAGAGTAA 2456

RESULT 24
AR344448
LOCUS AR344448 2531 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6582706.
ACCESSION AR344448
VERSION AR344448.1 GI:33740487
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 2531)
AUTHORS Johnson,L.S. and Adamou,J.E.
TITLE Vaccine compositions comprising Streptococcus pneumoniae
polypeptides having selected structural MOTIFS
JOURNAL Patent: US 6582706-A 11 24-JUN-2003;
FEATURES Location/Qualifiers
source 1..2531
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 41.5%; Score 991; DB 6; Length 2531;
Best Local Similarity 65.7%; Pred. No. 3.2e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
DB 60 TTCCTATGAGCTTGGACGTTTACCAAGCTGTGTCAGGATTAAGAAGAGTCTAATCGAGTTGC 119
QY 58 CTATATAGATCGAAAAACAGCGACGCCAAAAACCGAGAAATTTGACTCTGATGAGTTAG 117
DB 120 TTATATAGATGGTATGATCAGGCTGGTCAAAAGCGCAGAAAACTTGACACAGATGAAGTCAG 179
QY 118 CAAGGCTGAAGGAATCAATGCTGACCAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
DB 180 TAAGAGGAGGGGATCAACGCGCGAACAATTTGTTATCAAGATTACGGATCAAGGTTATGT 239
QY 178 CACTTTCACATGGCGACCACTATCATTTATCAATGGTAGGTTCCCTTATGACGCTATCAT 237
DB 240 GACCTCTCATGGAGACCATTAATCAATTAATGCAAGGTTCCCTTAATGATGCCATCAT 299
QY 238 CAGTGAAGAAATTCTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAA 297
DB 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATCAGTTGAAGGATCAGACATTTGTCAA 359
QY 298 TGAGGTCAGGCTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGA 357
DB 360 TGAATCAAGGGTGGTTATGTCATTAAGGTAACCGTAAATACATATGTTTACCTTAAGGA 419

```

```

358 TGCTGCCACGCGGATAACGTCGGTACAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
DB
420 TGCRCGCTCATGCGGATAAATATTCGACAAAAGAGAGATTAAACGTCAGAGCAGGAACG 479
QY
418 TAGTCAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGTTCCTTGGCAGCTTC 477
DB
480 CAGTCAATAATCAT-----AACTCAAGAGCAGATAATGCTGTGCTCGACGACAGC 530
QY
478 GCAGGACGCTATATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
DB
531 CCAGGACGTTTATCAACGAGATGATGGGTATATCTTCAATGCATCTGATATCATTAGGA 590
QY
538 TACTGTGTGATGCTTATATCGTCTCATGAGATCATTAACATTAATCTTCAAGATGA 597
DB
591 CACGGGTGATGCTTATATCGTCTCACGGCGACCATTAACATTAATCTTCAAGATGA 650
QY
598 GTTATCAGCTAGGAGTTGGCTGTCAGAGACCTTCTCTATCTGGTCGAGGAATCTGTC 657
DB
651 GTTATCAGCTAGGAGTTAGTCTGTCAGAGGCTTATGGAAATGG----- 695
QY
658 AAATTTCAAGAACCTTATCGCGACAAAATAGCGATAACACTTTCAAGAAACAACTGGGTACC 717
DB
696 -GAGACAGGATCTCGTCTCTTCTTCAAGTTCTAGTTTATAATGCAATCCAGCTCAACAA 754
QY
718 TTCTGTAAAGCAATCCAGGAACCTACAAATATCAACAAAGCAACAGACACATTAACAG 777
DB
755 GATTGTCAAGAACCAACAATCT-----GACTGTCACTCCAACTTA 794
QY
778 TCAAGCAAGTCAAGTAAATGACATGATAGTCTCTTGAAACAGCTCTTCAAACTGCTTT 837
DB
795 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAACCCTT 854
QY
838 GAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATFCAAGTCG 897
DB
855 ATCAAGACGCACTGTGAATCTGATGGCTTATTTTCGACCCAGCGCAATFCAAGTCG 914
QY
898 AACAGCTAGAGGTTTGGAGTCCACAGGAGATCATTAACACTTATCCCTTACTCTCA 957
DB
915 AACCGCAGAGGTGTAGCTGTCCCTCATGTGTAACCATTTACCACTTTATCCCTTATGAACA 974
QY
958 AATGCTCTGAATTTGAAGAACGANTCGTCTGATTTATTTCCCTTCTGTTATCTTCAAAACA 1017
DB
975 AATGCTCTGAATTTGAAGAACGANTGCTCTGATTTATTTCCCTTCTGTTATCTGTTCAACCA 1034
QY
1018 TTGGGTACCGAATTCAGAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
DB
1035 TTGGGTACCGAATTCAGACCCAGAACCAAGTCCACACCGACTCCAGAACCTAGTCC 1094
QY
1078 AGGCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCT 1137
DB
1095 AAGTCCGCAACCACTCCAGCAATCCAAATGTATGGGAA-----ATTGGTCAAGAAAGC 1148
QY
1138 GGTACGAAAAGTTCGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTCGTTATGCTTT 1197
DB
1149 TGTTCGAAAAGTAGCGATGTTATGCTTTTGAGGAGNATGAGTTTCTCGTTATATCCC 1208
QY
1198 TCGAAAAGATTTACCATCTGAAACTGTTTAAATACTTTGAAAAGCAAGTTATCAAAAAGA 1257
DB
1209 AGCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATTCATAGCAAACTGGCGCAGCAGA 1268
QY
1258 GAGTGTTCACACACTTTAACTGCTTAAAAAAGAAAATGTTGCTCTCTCGTACCAAGAAAT 1317
DB
1269 AAGTTTATCTCATAGCTAGGAACCTTAAGAAAACCTGACCTCCCATCTAGTATCGGAAT 1328
QY
1318 TTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGGTTGNAATAAAGG 1377
DB
1329 TTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGTATAATAAGG 1388
QY
1378 TCGTAATTTCTGATTTCAAGCCTTAGACAAATATTAGAACGCTTGATGATGATCGAC 1437
DB
1389 TCGCAAGATTTGAGGCTTTGGATAACCTGTTGGAACGACCTCAAGGATGCTCAAG 1448
QY
1438 TAAATAAGAAAATTTGGTAGATGATTTATTGGCAATTCCTGACCAACATTTACCATCCAGA 1497

```


[illegible]

RESULT 25			
AF318955			
LOCUS	2517 bp	DNA	linear
DEFINITION	Streptococcus pneumoniae pneumococcal histidine triad protein D precursor (phtD) gene, partial cds.		

ACCESSION AF318955.1 GI:12744743
 VERSION
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 2517)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T.,
 Dormitzer, M., Dagan, R., Brewah, Y.A., Barren, P., Lathigra, R.,
 Langermann, S., Koenig, S. and Johnson, S.
 Identification and characterization of a novel family of
 pneumococcal proteins that are protective against sepsis
 Infect. Immun. 69 (2), 949-958 (2001)
 21101045
 PUBMED
 11159990
 2 (bases 1 to 2517)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M. and
 Johnson, S.
 Direct Submission
 Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
 West Watkins Mill Road, Gaithersburg, MD 20878, USA
 LOCATION/Qualifiers
 1..2517
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="4"
 /db_xref="taxon:1313"
 1..>2517
 /genes="phtD"
 1..>2517
 /gene="phtD"
 /note="PhtD; protective pneumococcal antigen"
 /codon_start=1
 /transl_table=11
 /product="pneumococcal histidine triad protein D
 precursor"
 /protein_id="AAK06760.1"
 /db_xref="GI:12744744"
 /translation="MKINKYLAGSVAVIALSVCSVELGRHQAGQVKESNRVSVDIGD
 DAQGAENLTPDEVSREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIIE
 ELLMDPNQLKDSIVNIGKVIKDGKTYVYLKDAAHADNITKEEIKRQKEH
 SHNGGSDNDAVAAARAGQRYTTDDGYIFNASDIIEDTDGAYIVPHGDHYHVPKNE
 LSASELAARAAVWGAQGRSPSSSYNANPAQLSENHNLTVTPTYHONOGENTSS
 LARELVAPLSRHHVESDGLIFDPAQITSETARGVAPVHGNHYHFIPEQMSLEKRI
 ARILPLRVSNHWVDSRPEQSPQSTPEPSPQAPNPQAPSNFIDKULVKEAVR
 KYGDGVFEENGVSRIIPAKOLUAEATAGISKLAKQESLSHKLGAKKTDLPSSREF
 YNKAUDDLARIHQDLNKGQVDFEALNDLERIKDVPESDKVLGAKKTDLPSSREF
 PERLQKNPAQITTYTDEITQVAKLAGYTTEDGYIFDPRITDSEGDAYVTPMTHGSH
 IKKDSLSAEARAAQAAYAKEKGLTPSTPDQSGNTEAKGARAIYNRVKAAPVLDH
 MPNYLQYTVENKNGSLIIPHYDHYHNKIFEMFDEGLYEAPKQYTTLEDLLATVYVYEH
 PNERPHSDNGFGNASDHYVRKNKVDQSDKPEDKEHDEUSEPHSPSDEKENHAGLNPS
 ADNLKYPSDTTETEEEAEDTDEAIPQVENSVINAKIADAELAEKVTDPDSIRQNA
 METLTGLKSLSLLLGTGKNDNTTISAEDVSLALLKESQPAPIQ"
 1..87
 /gene="phtD"
 247..264
 /gene="phtD"
 /note="Region: histidine triad"
 418..495
 /gene="phtD"
 /note="Region: coiled-coil domain"
 619..636
 /gene="phtD"
 /note="Region: histidine triad"
 943..960
 /gene="phtD"
 /note="Region: histidine triad"
 1417..1509
 /gene="phtD"
 /note="Region: coiled-coil domain"
 1678..1695
 /gene="phtD"

misc_feature		/note="Region: histidine triad"		1930..1947	
repeat_region		/gene="phdD"		/note="Region: histidine triad"	
repeat_region		526..717		/rpt_type=direct	
repeat_region		1597..1779		/rpt_type=direct	
ORIGIN					
Query Match 41.4%; Score 990; DB 1; Length 2517;					
Best Local Similarity 68.1%; Pred. No. 5.2e-201;					
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;					
Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAA--	TAATCGTGTTC	57	
Db	60	TTCTATGAACTTGGTCGTCAACCAAGCTGGCTAAGAAAGAGTCTAATCGAGTTTC	119		
Qy	58	CTATATAGATGGAAACAAGCGACGCAAAAAACGGAGAAATTGACTCTGATGAGGTTAG	117		
Db	120	TTATATAGATGGTATCAGGCTGTCAAAAGCGAGAAACTTGACACCAAGATGAAGTCAG	179		
Qy	118	CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT	177		
Db	180	TAAGAGGGAGGGATCAACGCCGAAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT	239		
Qy	178	CACHTCACATGGCGACCACTATCATATTATTAAGTGAAGTTCTTATGACGCTATCAT	237		
Db	240	GACCTCTCATGGAGACCATTAATCAATTAATTAAGTGAAGTTCTTATGATGCGCATCAT	299		
Qy	238	CAGTGAAGAAATTACTCATGAAGATCAAACTATAAGCTAAAAGATGAGGATATTTGTA	297		
Db	300	CAGTGAAGACTCTCATGAAGATCCGAATTATCAGTTGAAGGATTGAGACATTGTCAA	359		
Qy	298	TCAGGTTCAAGGTTGGATATGTTATCAAGTATGATGGAATAATCTATGTTTACCTTAGGA	357		
Db	360	TGAATCAAGGTTGGTATGTTATCAAGGTATGAGAAATAATCTATGTTTACCTTAGGA	419		
Qy	358	TGCTGCCACCGGATAACGTCCTGACAAAGAGGAATCAATCGACAAAAACAAGACA	417		
Db	420	TGCAGCTCATCGGATATATTCGGACAAAGAGAGATTAAAGTCAAGAGCAGAGACA	479		
Qy	418	TAGTCAACATCGTGAAGTGGAACTCCAAGAAACGATGGTGTGTCCTTGGCACGTTTC	477		
Db	480	CAGTCAATATCAACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGACGACAGAGC	533		
Qy	478	GCAAGGACGCTATACACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGA	537		
Db	534	CCAAGGACGCTATACCAAGGATGATGTTATATCTTCAATGCAATCTGATATCATTTGAGA	593		
Qy	538	TACTGGTATGCTTATATCGTTCCTCATGGAGATCAATTACCATTAACATTCTTAAGAATGA	597		
Db	594	CACGGGTGATGCTTATATCGTTCCTCAAGCGACCATTAACATTACATTCTTAAGAATGA	653		
Qy	598	GTTATCAGCTAGCGAGTTGGTGTGTCAGAAAGCCCTTCCCTATCTGGTCGAGGAATCTGTC	657		
Db	654	GTTATCAGCTAGCGAGTTAGTGTGTCGAGAAGCCT-----	688		
Qy	658	AAATTCAGAACCTTATCCGCGACAAATATACCGGATAACACTTCAAGAAACAACTGGGTACC	717		
Db	689	-----ATTGGAATGGGAACGAGGATCTCGTCTCTTCAAGTTCTAGTTATAATGC	740		
Qy	718	TTCTGTGAAGCAATCCAGAACTACAAATATCTAACAGACAAACACCTTAACAG	777		
Db	741	AAATCCAGCTCAACCAAGATTGTGAGAGAACCACAAATCTGACTGTCACTCCAACTTATCA	800		
Qy	778	TCAAGCAAGTCAAGTAATGACATTTGATGCTCTTTGAACAGCTCTTCAAACTGCCCTTT	837		
Db	801	TCA---AAATCAAGGGGAAAAATTTCAAGCCCTTTTACGTGAATTTGATGCTTAACCCCTT	857		
Qy	838	GAGTCAACGACATGATGATCTGATGGCCTTGTCTTTGATPCCAGCAACAAATCACAAGTCG	897		
Db	858	ATCAGAACGCCATGTGGAATCTGATGGCCCTTATTTTCGACCCGCGCAATCAAGTCG	917		

Qy	898	AACAGCTAGAGGTGTTGCAGTCCACACGAGATCAATTACCACCTTCATCCCTTACTCTCA	957
Db	918	AACCGCCAGAGGTGTAGTGTCCCTCATGTGTAAACCACTTACCACTTATCCCTTATGAACA	977
Qy	958	AATGCTGAATTGGAAGAACGAATCGCTCGTATATATCCCTTCGTTATCGTTTCAACACCA	1017
Db	978	AATGCTGAATTGGAAGAACGAATCGCTCGTATATATCCCTTCGTTATCGTTTCAACACCA	1037
Qy	1018	TTGGGTACAGATTCAAGGCCAGAACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC	1077
Db	1038	TTGGGTACAGATTCAAGGCCAGAACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC	1097
Qy	1078	AGCCCGCAACCTGCAACCAATCTTAAATATAGACTCAAAATCTTC-----TTT	1125
Db	1098	AAGTCCGCAACCTGCAACCAATCTTAAATATAGACTCAAAATCTTC-----TTT	1157
Qy	1126	GGTTAGTCAGCTGGTACGAAAGAGTTGGGGAAGATATGTTATCGAAGAAAGGCACTC	1185
Db	1158	GGTCAAGAAAGCTGTTTCGAAAGTAGGCGATGTTATGTTCTTTGAGGAGATGAGGTTTC	1217
Qy	1186	TCGTTATGCTTTTCGAAAGATTTACCATCTGAAACTGTTTAAAAATCTTGAAACCAAGTT	1245
Db	1218	TCGTTATATCCGACCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACT	1277
Qy	1246	ATCAAAAACAAGAGAGTGTTCACACACTTTTAACTGCTTAAAAAAGAAATGTTGCTCTCG	1305
Db	1278	GGCCAAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCCATCTAG	1337
Qy	1306	TGACCAAGAAATTTATGATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGT	1365
Db	1338	TGATCGAGAAATTTTACAAATGAGGCTTATGACTTACTAGCAGAAATTCACCAAGATTTACT	1397
Qy	1366	TGNAATTAAGGCTCGTAAATTTCTGATTTCCAAAGCTTAGACAAATTTATAGAACCTTGAA	1425
Db	1398	TGATAATTAAGTTCGACAGAGTTGATTTTGGCTTTTCGATAACCTGTTGGAACGACTCAA	1457
Qy	1426	TGATGAATTCGACTAATTAAGAAATAATTTGGTAGATGATTTTATTTGGCATCTCTAGCACCAAT	1485
Db	1458	GGATGTCCCAAGTGTATAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGTCCGAT	1517
Qy	1486	TACCATTCGAGCGGACTTGCACAAACCAAAATTCCTCAAAATGAGTATCTAGAAACGAGAT	1545
Db	1518	TCGTCATTCGAGAGCTTTAGGAAACCAAAATTCGCAAAATTTACCTACACTGATGATGAT	1577
Qy	1546	TCGTTATTCCTCAATTAGCTGATAAGTATACAACTGAGATGGTTTACATTTTGTATGAACA	1605
Db	1578	TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACAGAGAGCGTTTATATCTTTGATCCCTCG	1637
Qy	1606	TGATAATATCAGTGTGATGAAGGAGATGCAATATGTAAACCGCTCATATGGGCCATAGTCACTG	1665
Db	1638	TGATAATACCACTGATGAGGGGATGCTTATGTAATCTCAATATGACCCATAGCCACTG	1697
Qy	1666	GATTGGAAAAAGATAGCCCTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTACTATAAGA	1725
Db	1698	GATTAAAAAGATAGTTTGTCTGAGCTGAGAGAGCGGACCCAGGCTTATGTAAGA	1757
Qy	1726	AAAAAGGTATCTTACTCTCCATCTCCAGACGAGATGTTTAAAGCAAAATCAACTGAGAGATG	1785
Db	1758	GAAAGGTTTACCCCTCTCTTCGACAGACCATCAGGATTCAGGAAATATCTAGGCAAAAAGG	1817
Qy	1786	TGCAGAGCTATTTTACAAATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCGACTTCC	1845
Db	1818	AGCAGAGCTATCTACACCGCTGAAAGAGAGCTTAAGAAAGTGGCACTTGATCGTATGCC	1877
Qy	1846	ATATATGTTTCAGGATACAGTTGAGTTTAAAAACGTTTAAATTTGATTTTCTCTATAAGGA	1905
Db	1878	TTACAACTCTCAATATCTGTAGAGTCAAAAACGTTAGTTTATCATCACTCATATGA	1937
Qy	1906	TCATTACCAATATTTAAATTTGCTTGTGTTGATGATCACATACAAAGCTTCAAAATGG	1965
Db	1938	CCATTACCAATCAATCAAAATTTGAGTGGTTTGAACGAAAGCCCTTTATAGGCACTTAAAGG	1997

QY	1966	CTATACCTTGGAGATTGTTTGGACGATTAAAGTACTACGTAGACACACCCCTGACGAACG	2025
Db	1998	GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACG	2057
QY	2026	TCCACATTCTAATGATGGATGGGGCAATGCCAGTGCAGTCATGTGTTAGGCAAGAAAGACCA	2085
Db	2058	TCCGCATTTCAGATAAATGGTTTGGTAAAGCTAGCGACCATGTTTCGTAAAAAATAAGGTAGA	2117
QY	2086	CAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC	2145
Db	2118	CCAGACAGTAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACTCACCC	2177
QY	2146	TGAGCCAGAAG	2156
Db	2178	TGAATCTGATG	2188

Search completed: September 30, 2004, 18:25:56
Job time : 9320 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 11:00:45 ; Search time 915 Seconds
(without alignments)
11091.743 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	2	AAV27351 Streptococ
2	2388	100.0	2389	6	ABQ84819 S. pneumo
3	2388	100.0	2389	9	ADC45136 Streptococ
4	2388	100.0	2406	7	ABX06886 S. pneumo
5	2388	100.0	2451	3	AAA47604 Streptococ
6	2388	100.0	8195	2	AAV52227 Streptococ
7	2388	100.0	110000	7	ABSS56454_09 Continuation (10 o
8	2388	100.0	110000	7	ABSS56454_10 Continuation (11 o
9	1374.2	57.5	2523	3	AAA65736 Streptococ
10	1374.2	57.5	2647	3	AAA65736 Streptococ
11	1374.2	57.5	2647	6	ABK15103 DNA encod
12	1001.8	42.0	2478	3	AAA08557 S. pneumo
13	1003.8	42.0	2481	3	AAA05417 Streptococ
14	993	41.6	2457	7	ABX06885 S. pneumo
15	991	41.5	2531	3	AAA47605 Streptococ
16	990	41.4	2517	7	ABX06705 Streptococ
17	990	41.4	2531	3	AAA47602 Streptococ
18	990	41.4	110000	7	ABSS56454_08 Continuation (9 of
19	987.6	41.3	2290	2	AAV27356 Streptococ
20	987.6	41.3	2290	6	ABQ84824 S. pneumo
21	987.6	41.3	2290	9	ADC45146 S. pneumo
22	980.8	41.1	2639	3	AAA65737 Streptococ
23	980.8	41.1	2639	6	ABK15104 DNA encod

24	859.4	36.4	2163	2	AAV25394 Streptococ
25	653.8	27.4	2359	2	AAV52376 Streptococ
26	548.4	23.0	3171	3	AAA65739 Streptococ
27	385.4	16.1	1342	2	AAV27414 Streptococ
28	385.4	16.1	1342	6	ABQ84882 S. pneumo
29	385.4	16.1	1342	9	ADC45262 S. pneumo
30	385.4	16.1	1455	3	AAA05473 Streptococ
31	385.4	16.1	1455	3	AAZ91804 Streptococ
32	385.4	16.1	1455	3	AAA65733 Streptococ
33	385.4	16.1	1455	3	AAA47603 Streptococ
34	385.4	16.1	3117	7	ABX06706 Streptococ
35	385.4	16.1	3120	3	AAA65730 Streptococ
36	385.4	16.1	5048	3	AAA65735 Streptococ
37	385.4	16.1	5048	6	ABK15101 DNA encod
38	385.4	16.1	6867	2	AAV52325 Streptococ
39	381.4	16.0	973	2	AAV52488 Streptococ
40	347	14.5	2528	6	AAA65738 Streptococ
41	347	14.5	2528	6	ABK15105 DNA encod
42	247.4	10.4	504	3	AAA08556 S. pneumo
43	243.4	10.2	1684	2	AAV52391 Streptococ
44	214.6	9.0	2475	6	ABN66838 Streptococ
45	214.6	9.0	2478	5	AAV52391 Streptococ
46	211.4	8.8	1146	3	AAA00036 Streptococ
47	211.4	8.8	2466	6	ABN69535 Streptococ
48	211.4	8.8	2466	6	ABN70334 Streptococ
49	211.4	8.8	2469	3	AAA05811 Streptococ
50	211.4	8.8	2469	5	AAV52325 Streptococ
51	211.4	8.8	110000	6	ABN71527_12 Continuation (13 o
52	209.8	8.8	2469	3	AAA65740 Streptococ
53	209.8	8.8	5215	2	AAV52391 Streptococ
54	206.6	8.6	2472	3	AAA65741 Streptococ
55	206	8.6	110000	6	ABN71527_18 Continuation (19 o
56	201.6	8.4	1398	6	ABN66839 Streptococ
57	201.2	8.4	816	8	ACF42878 S. pneumo
58	201.2	8.4	816	8	ACF42878 S. pneumo
59	201.2	8.4	819	6	ABK15106 DNA encod
60	201.2	8.4	819	8	ACF42881 S. pneumo
61	200.8	8.4	2721	8	ACF42880 S. pneumo
62	188.2	7.9	492	2	AAV42952 Streptococ
63	117.4	4.9	1455	3	AAA05838 Streptococ
64	109	4.6	841	2	AAT28529 S. pneumo
65	109	4.6	841	4	ABA76858 Streptococ
66	102.6	4.3	1137	6	ABN66837 Streptococ
67	62.8	2.6	2000	7	ADA71938 Rice gene
68	53.2	2.2	1729	2	AAV42952 Streptococ
69	52.8	2.2	1168	9	ADC87372 Human GPC
70	51.8	2.2	567	7	ABX06707 S. pneumo
71	51.8	2.2	1083	5	AAV576745 DNA encod
72	51.6	2.2	3060	4	AAH54131 S. epider
73	51.6	2.2	3309	4	AAH54131 S. epider
74	51.6	2.2	11076	7	ACA47181 Prokaryot
75	51.6	2.2	11091	6	ABN92780 Staphyloc
76	51.2	2.1	15935	4	ABL06114 Streptococ
77	50	2.1	7662	6	ABN80080 Human cha
78	49.8	2.1	1959	4	AAI14146 Human fo
79	49.8	2.1	1959	4	AAH5872 Human fo
80	49.8	2.1	1959	4	AAI35531 Human bre
81	49.8	2.1	1959	4	ABA45382 Human bre
82	49.8	2.1	1959	4	ABA25546 Probe #40
83	49.8	2.1	1959	4	AAK29571 Human bon
84	49.8	2.1	1959	4	AAK04090 Human bra
85	49.8	2.1	1959	4	ABS29202 Human liv
86	49.8	2.1	1959	5	AAI03994 Probe #39
87	49.8	2.1	1959	6	ABS04127 Human gen
88	48.6	2.0	3399	2	AAT05868 Chicken 1
89	48.2	2.0	3579	3	AAV70099 Plasmodu
90	48	2.0	2631	6	ABN68134 Streptococ

ALIGNMENTS

AAV27351	DB	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	180
ID AAV27351 standard; DNA; 2389 BP.	QY	181	TTCCATGCGGACCACTATCAATTAATGTAAGGTTCCCTATGACGCTATCATCAG	240
AC AAV27351;	DB	181	TTCCATGCGGACCACTATCAATTAATGTAAGGTTCCCTATGACGCTATCATCAG	240
XX	QY	241	TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGATATGTTAATGA	300
DT 02-OCT-1998 (first entry)	DB	241	TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGATATGTTAATGA	300
Streptococcus pneumoniae SP0036 nucleotide.	QY	301	GGTCAAGGTTGGATATGTTATCAAGGTAGTGAAGAAATCAATGCTTAAAGGATGC	360
Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;	DB	301	GGTCAAGGTTGGATATGTTATCAAGGTAGTGAAGAAATCAATGCTTAAAGGATGC	360
Streptococcus pneumoniae; otitis media; meningitis; ss.	QY	361	TGCCCCGCGGATAACGTCGTATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
Streptococcus pneumoniae.	DB	361	TGCCCCGCGGATAACGTCGTATCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
Key Location/Qualifiers	QY	421	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTCTGCTTGGCAGCTTCGCA	480
2. 2389	DB	421	TCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGTCTGCTTGGCAGCTTCGCA	480
/*tag= a	QY	481	AGGACGCTATACAGATGATGTTTATATCTTTAATGCTTCTCATATCATAGAGGATAC	540
/product= "SP0036"	DB	481	AGGACGCTATACAGATGATGTTTATATCTTTAATGCTTCTCATATCATAGAGGATAC	540
/transl_except= (pos.1367..1369,aa:Xaa)	QY	541	TGGTGATGCTTATATCGTCTCATGAGATCAATACCATTAATCAATCCTAAGAAATGATT	600
/note= "no stop codon given; Xaa is unspecified"	DB	541	TGGTGATGCTTATATCGTCTCATGAGATCAATACCATTAATCAATCCTAAGAAATGATT	600
WO9818930-A2.	QY	601	ATCAGTAGCGAGTTGCTGTCAGAGGCTTCTATCTGCTGAGGAAATCTGTCAAA	660
07-MAY-1998.	DB	601	ATCAGTAGCGAGTTGCTGTCAGAGGCTTCTATCTGCTGAGGAAATCTGTCAAA	660
30-OCT-1997; 97WO-US019422.	QY	661	TTCAAGAACCTATCGCGACAAATAGCGATAACACTTCAAGAAACAACTGGTACCTTC	720
31-OCT-1996; 96US-0029960P.	DB	661	TTCAAGAACCTATCGCGACAAATAGCGATAACACTTCAAGAAACAACTGGTACCTTC	720
(HUMA-) HUMAN GENOME SCI INC.	QY	721	TGTAAGCAATCCAGGAACTACAAATATCTAAACAGCAACCAAGCAACCTAAGCTCA	780
Kunsch CA, Choi GH, Johnson LS, Hromockyj A;	DB	721	TGTAAGCAATCCAGGAACTACAAATATCTAAACAGCAACCAAGCAACCTAAGCTCA	780
WPI; 1998-272224/24.	QY	781	AGCAAGTCAAAAGTAAATGATGATGCTCTTGAACAGCTCTACAACTGCTTTCAG	840
P-PSDB; AAW55090.	DB	781	AGCAAGTCAAAAGTAAATGATGATGCTCTTGAACAGCTCTACAACTGCTTTCAG	840
Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae	QY	841	TCAACGATCATGATGATGCTCTTGAACAGCTCTACAACTGCTTTCAG	900
- or their epitope-containing fragments, useful in protective or	DB	841	TCAACGATCATGATGATGCTCTTGAACAGCTCTACAACTGCTTTCAG	900
therapeutic vaccines, and for diagnosis.	QY	901	AGCTAGAGTGTGTCAGTCCACACGAGATCAATACCATTCATCCCTTACTCTCAAT	960
Claim 1; Page 59; 118pp; English.	DB	901	AGCTAGAGTGTGTCAGTCCACACGAGATCAATACCATTCATCCCTTACTCTCAAT	960
The present sequence encodes a protein from Streptococcus pneumoniae. The	QY	961	GTCGAAATGGAAGAACGAAATCGCTGATATATATCCCTTCTGTTATCGTTCAAACCATG	1020
nucleic acid sequence encoding the Streptococcus pneumoniae protein can	DB	961	GTCGAAATGGAAGAACGAAATCGCTGATATATATCCCTTCTGTTATCGTTCAAACCATG	1020
be useful in vaccines for inducing protective antibodies against	QY	1021	GGTACCATGATCAAGGCGAGAACCAACCAAGTCCACACGACTCCGGAACCTAGTCCAGG	1080
Streptococcus pneumoniae, for treatment or prevention of infection e.g.	DB	1021	GGTACCATGATCAAGGCGAGAACCAACCAAGTCCACACGACTCCGGAACCTAGTCCAGG	1080
pneumonia, otitis media or meningitis. Probes based on the nucleic acid	QY	1081	CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTGTAGTCACTGGT	1140
are used to detect Streptococcus infection (by usual hybridisation or	DB	1081	CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTGTAGTCACTGGT	1140
amplification methods), also for isolating Streptococcus genes or their	QY	1141	AGGAAAGTTGGGAGGATATGATTCGAGAAAGGAGGATCTCTCGTTATGCTTTTCG	1200
allelic variants. The protein can be used similarly to detect specific	DB	1141	AGGAAAGTTGGGAGGATATGATTCGAGAAAGGAGGATCTCTCGTTATGCTTTTCG	1200
antibodies in standard immunoassays, especially for diagnosing or	QY	1201	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTATCAAAACAGAGAG	1260
monitoring infections. Antibodies which bind the protein are used to	DB	1201	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTATCAAAACAGAGAG	1260
detect corresponding antigens, to purify the protein and for passive	QY			
immunisation (optionally coupled to a toxin). Vaccines are administered,	DB			
e.g. by injection, orally or through the skin, typically at 0.01-1000	QY			
(especially 10-300) mu g/ml per dose	DB			
Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;	QY			
Query Match 100.0%; Score 2388; DB 2; Length 2389;	DB			
Best Local Similarity 100.0%; Pred. No. 0;	DB			
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	1	TTCTACGAGTTGGGACTGTATCAAGCTAGACGTTAAGGAAATATCGTTTCCTA	60
	DB	1	TTCTACGAGTTGGGACTGTATCAAGCTAGACGTTAAGGAAATATCGTTTCCTA	60
	QY	61	TATAGATGAAACAAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120
	DB	61	TATAGATGAAACAAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120
	QY	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	180

1261 TGTTTCCACACATTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGAACAAGATTTTA 1320
 1261 TGTTTCCACACATTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGAACAAGATTTTA 1320
 1321 TGATAAGCATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGCTCG 1380
 1321 TGATAAGCATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGCTCG 1380
 1381 TAAATCTGATTTTCCAGCCTTAGACAAATTTAGAACGTTGATGATGAATCGACTAA 1440
 1381 TAAATCTGATTTTCCAGCCTTAGACAAATTTAGAACGTTGATGATGAATCGACTAA 1440
 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCAGCG 1500
 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTTACCCATCCAGCG 1500
 1501 ACTTGGCAAAACCAATTTCTCAAAATTTGAGTATCTGAAGACGAAATTCGTATTGCTCAAT 1560
 1501 ACTTGGCAAAACCAATTTCTCAAAATTTGAGTATCTGAAGACGAAATTCGTATTGCTCAAT 1560
 1561 AGCTGATAGTATACAACTGATGATTTTACATTTTGTGATGAACATGATATATCATGTA 1620
 1561 AGCTGATAGTATACAACTGATGATTTTACATTTTGTGATGAACATGATATATCATGTA 1620
 1621 TGAAGGAGATGATATGTAACGCCCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680
 1621 TGAAGGAGATGATATGTAACGCCCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680
 1681 CTTTCTGATAGGAAAAAGTTGCGAGCTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1740
 1681 CTTTCTGATAGGAAAAAGTTGCGAGCTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1740
 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGGAGAGCTATTTA 1800
 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGGAGAGCTATTTA 1800
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCACATTCATATATGTTGAGCA 1860
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCACATTCATATATGTTGAGCA 1860
 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTAACATAATAT 1920
 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTAACATAATAT 1920
 1921 TAAATTTGCTTTGTTGATGATCACAATCAAACTGATGATGATGATGATGATGATGATG 1980
 1921 TAAATTTGCTTTGTTGATGATCACAATCAAACTGATGATGATGATGATGATGATGATG 1980
 1981 TTTGTTTGGCAGGATTAAGTACTAGTATAGAACACCCCTGACGAACCTCCACATTTCTAATGA 2040
 1981 TTTGTTTGGCAGGATTAAGTACTAGTATAGAACACCCCTGACGAACCTCCACATTTCTAATGA 2040
 2041 TGGATGGGGCAATGCGAGTGCATGTTTGGGCAAGAAAGACCACTGAGATGATGATGATG 2100
 2041 TGGATGGGGCAATGCGAGTGCATGTTTGGGCAAGAAAGACCACTGAGATGATGATGATG 2100
 2101 TAAGAACTTCAAGCGGATGAGGCGCAGTATAGGAAACACCTGCTGAGCAGAGTCTCC 2160
 2101 TAAGAACTTCAAGCGGATGAGGCGCAGTATAGGAAACACCTGCTGAGCAGAGTCTCC 2160
 2161 TCAAGTATGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAATTTGCTTGGCAA 2220
 2161 TCAAGTATGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAATTTGCTTGGCAA 2220
 2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAAATCTAGCTGTTTACGAAATAA 2280
 2221 AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAAATCTAGCTGTTTACGAAATAA 2280
 2281 TTTGACTCTTCAAAATTTAGTATACATAGTATCATGSCAGAGCAAGAAATTTACTTGC 2340
 2281 TTTGACTCTTCAAAATTTAGTATACATAGTATCATGSCAGAGCAAGAAATTTACTTGC 2340

QY 2341 GTTGTAAAAGGAAGTAATCCTTCATCTCTTAAGTAAGGAAAAAATAAAC 2389
 Db 2341 GTTGTAAAAGGAAGTAATCCTTCATCTCTTAAGTAAGGAAAAAATAAAC 2389
 RESULT 2
 ABQ84819
 ID ABQ84819 standard; DNA; 2389 BP.
 XX ABQ84819;
 AC AC
 XX 04-SEP-2002 (first entry)
 XX S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 XX Streptococcus pneumoniae; Streptococcal infection; detection; gene; ds.
 KW Streptococcus pneumoniae.
 OS US2002061545-A1.
 PN 23-MAY-2002.
 XX 22-JAN-2001; 2001US-00765272.
 PF 30-OCT-1997; 97US-00961083.
 XX (CHOI/) KUNSH C A.
 PA (KUNSH/) KUNSH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI; 2002-479261/51.
 DR P-PSDB; ABP54584.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX Claim 1; Page 27; 70pp; English.
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;
 SQ
 Query Match 100.0%; Score 2389; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 60
 Db 1 TTCTTACAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 60
 QY 61 TATAGATGGAACCAACGACGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 Db 61 TATAGATGGAACCAACGACGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAA 120
 QY 121 GCGTGAAGGATCAATCTGCTGAGCAATCGTCATCAAGATAACAGACCAAGCTATGTAC 180

QY 2341 GTTGTAAAAGGAAGTAATCTTCATCTGTAAAGTAAGGAAAAATAAAC 2389
 |||||
 Db 2341 GTTGTAAAAGGAAGTAATCTTCATCTGTAAAGTAAGGAAAAATAAAC 2389

RESULT 3
 ID ADC45136
 ADC45136 standard; DNA; 2389 BP.
 XX
 AC ADC45136;
 XX
 DT 18-DEC-2003 (first entry)
 XX

DE S. pneumoniae DNA encoding antigen SP036.
 XX
 KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US6573082-B1.
 XX

XX 03-JUN-2003.
 XX

XX 28-MAR-2000; 2000US-00536784.
 XX

XX 31-OCT-1996; 96US-0029960P.
 PR

XX 30-OCT-1997; 97US-00961083.
 PR

XX (HUMA-) HUMAN GENOME SCI INC.
 XX

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Rannon MR;
 PI Rosen CA;
 PI

XX WPI: 2003-764574/72.
 DR

XX P-PSDB; ADC45137.
 DR

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by Streptococcus pneumoniae.
 PT

PS Example 1; SEQ ID NO 55; 58pp; English.
 PS

CC The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence encodes
 CC an S. pneumoniae antigenic protein.
 XX

SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Query Match 100.0%; Score 2388; DB 9; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA 60
 |||||
 Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA 60

QY 61 TATAGATGAAACAAAGCGACCAAAAAACCGAGAAATTGACTCCTGATGAGGTTAGCAA 120
 |||||

Db 61 TATAGATGAAACAAAGCGACCAAAAAACCGAGAAATTGACTCCTGATGAGGTTAGCAA 120
 |||||

QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 180
 |||||

Db 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTAC 180
 |||||

QY 181 TTCAATGGCGACCACTATCATTTATCAATGGTAAGTTCTTATGACGCTATCATCAG 240
 |||||
 Db 181 TTCAATGGCGACCACTATCATTTATCAATGGTAAGTTCTTATGACGCTATCATCAG 240
 |||||
 QY 241 TGAAGAAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
 |||||
 Db 241 TGAAGAAATTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
 |||||
 QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
 |||||
 Db 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360
 |||||
 QY 361 TGCCACGCGGATAACGTCCTGTAACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
 |||||
 Db 361 TGCCACGCGGATAACGTCCTGTAACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
 |||||
 QY 421 TCAACATCGTGAAGGTGGAACTCCAAGAACGATGGTGTGCTTGGCTTGGCACGTTGCGCA 480
 |||||
 Db 421 TCAACATCGTGAAGGTGGAACTCCAAGAACGATGGTGTGCTTGGCTTGGCACGTTGCGCA 480
 |||||
 QY 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 |||||
 Db 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 |||||
 QY 541 TGGTGAATGCTTATATCGTTCCTCATGGAGATCATTAACCATTCCTTAAGAAATCAGTT 600
 |||||
 Db 541 TGGTGAATGCTTATATCGTTCCTCATGGAGATCATTAACCATTCCTTAAGAAATCAGTT 600
 |||||
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCGAAGCCTTCTCTGTCGAGGAAATCTGTCAA 660
 |||||
 Db 601 ATCAGCTAGCGAGTTGGCTGCTGCGAAGCCTTCTCTGTCGAGGAAATCTGTCAA 660
 |||||
 QY 661 TTCAAGAACCTATCGCCGCAAAAAATAGCGATAACCTTCAAGAAACAACTGGGTACTTTC 720
 |||||
 Db 661 TTCAAGAACCTATCGCCGCAAAAAATAGCGATAACCTTCAAGAAACAACTGGGTACTTTC 720
 |||||
 QY 721 TGTAGCAATCAGGAACTACAAAATACAAACAGCAACAAACAGCAACCACTAACAGTCA 780
 |||||
 Db 721 TGTAGCAATCAGGAACTACAAAATACAAACAGCAACAAACAGCAACCACTAACAGTCA 780
 |||||
 QY 781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAACAGCTCTACAACTGCCTTTGAG 840
 |||||
 Db 781 AGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAACAGCTCTACAACTGCCTTTGAG 840
 |||||
 QY 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCACAAATCACAAGTCGAAC 900
 |||||
 Db 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCACAAATCACAAGTCGAAC 900
 |||||
 QY 901 AGCTAGAGGTGTGCGAGTCCACACGAGATCATTAACCACTTCTCTTACTCTCAAT 960
 |||||
 Db 901 AGCTAGAGGTGTGCGAGTCCACACGAGATCATTAACCACTTCTCTTACTCTCAAT 960
 |||||
 QY 961 GTCTGAATTGGAAGAAACGAATCGCTCGTATTATTCCTTCTGTTATCGTTCAAACCATTTG 1020
 |||||
 Db 961 GTCTGAATTGGAAGAAACGAATCGCTCGTATTATTCCTTCTGTTATCGTTCAAACCATTTG 1020
 |||||
 QY 1021 GGTACCAAGATTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCCGAACCTAGTCCAGG 1080
 |||||
 Db 1021 GGTACCAAGATTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCCGAACCTAGTCCAGG 1080
 |||||
 QY 1081 CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCTAGCTGGT 1140
 |||||
 Db 1081 CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCTAGCTGGT 1140
 |||||
 QY 1141 ACGAAAAGTTGCGGAAGGATATGTTTCGAGAGAAAGGCAATCTCTGTTATGTTTTC 1200
 |||||
 Db 1141 ACGAAAAGTTGCGGAAGGATATGTTTCGAGAGAAAGGCAATCTCTGTTATGTTTTC 1200
 |||||
 QY 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTTAAAGCAAGTTCATCAAAACAAGAGAG 1260
 |||||
 Db 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTTAAAGCAAGTTCATCAAAACAAGAGAG 1260
 |||||

QY 1261 TGTTCACACACTTTAACTGCTTAAAAAGAAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
Db 1261 TGTTCACACACTTTAACTGCTTAAAAAGAAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
QY 1321 TGAATAAGGACATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTTCGNAAATAAGGCTCG 1380
Db 1321 TGAATAAGGACATATAAATCTGTTAACTGAGGCTCATAAAGCCTTGTTCGNAAATAAGGCTCG 1380
QY 1381 TAATCTGATTTTCAAGCCCTTAGACAAATATTAGAAACGCTTGAATGATGATCGACTAA 1440
Db 1381 TAATCTGATTTTCAAGCCCTTAGACAAATATTAGAAACGCTTGAATGATGATCGACTAA 1440
QY 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGACCAATACCCATCCAGAGCG 1500
Db 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGACCAATACCCATCCAGAGCG 1500
QY 1501 ACTTGGCAAAACCAATTTCTCAAAATTTAGTATCTATGAAAGCAAGTTTGGTATTTGCTCAATT 1560
Db 1501 ACTTGGCAAAACCAATTTCTCAAAATTTAGTATCTATGAAAGCAAGTTTGGTATTTGCTCAATT 1560
QY 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTATGAAATGATATATCACTGA 1620
Db 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTATGAAATGATATATCACTGA 1620
QY 1621 TGAAGGATGCTATATGTAAGCCCTCATATGGGCCATAGTCACGTGATTTGAAAGATAG 1680
Db 1621 TGAAGGATGCTATATGTAAGCCCTCATATGGGCCATAGTCACGTGATTTGAAAGATAG 1680
QY 1681 CCTTCTCTGATAAGGAAAAGTTGAGCTCAAGCCTTACTATAAGAAAAGGATTCCTTACC 1740
Db 1681 CCTTCTCTGATAAGGAAAAGTTGAGCTCAAGCCTTACTATAAGAAAAGGATTCCTTACC 1740
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATTCACCTGAGATAGTCAGCAGCTATTTA 1800
Db 1741 TCCATCTCCAGACGAGATGTTAAAGCAATTCACCTGAGATAGTCAGCAGCTATTTA 1800
QY 1801 CAATCGTGTGAAAGGAAAAGCAATTCACCTGCTTCCATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAAGGAAAAGCAATTCACCTGCTTCCATATATGTTGAGCA 1860
QY 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTAATTCCTCATAGGATCATACCATATAT 1920
Db 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTAATTCCTCATAGGATCATACCATATAT 1920
QY 1921 TAAATTTGCTGTTTGTATGATCACATACAAAGCTTCAATGCTATCTCTGGAGA 1980
Db 1921 TAAATTTGCTGTTTGTATGATCACATACAAAGCTTCAATGCTATCTCTGGAGA 1980
QY 1981 TTTGTTTCGACGATTAAGTACTAGCTAGAACACCCCTGACGAACTCCACATTTCTAATGA 2040
Db 1981 TTTGTTTCGACGATTAAGTACTAGCTAGAACACCCCTGACGAACTCCACATTTCTAATGA 2040
QY 2041 TGGATGGGGCAATGCACTGAGCATGTTGTTAGGCAAGAAAGACCAACGATGAGATCCAAA 2100
Db 2041 TGGATGGGGCAATGCACTGAGCATGTTGTTAGGCAAGAAAGACCAACGATGAGATCCAAA 2100
QY 2101 TAAGAACTTCAAGCGGATGAGACCGCTAGAGCAACACCTGCTGAGCAGAGTCCC 2160
Db 2101 TAAGAACTTCAAGCGGATGAGACCGCTAGAGCAACACCTGCTGAGCAGAGTCCC 2160
QY 2161 TCAAGTAGAGCTGAAAAAGTAAAGAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGCAA 2220
Db 2161 TCAAGTAGAGCTGAAAAAGTAAAGAGCCCACTCAAGAAAGCAGAAAGTTTCTTGGCAA 2220
QY 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATAA 2280
Db 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATAA 2280
QY 2281 TTTGACTCTTCAAAATTTGATAAACAATAGTATCATGGCAGAGCAGAAAAAATTTACTTGC 2340
Db 2281 TTTGACTCTTCAAAATTTGATAAACAATAGTATCATGGCAGAGCAGAAAAAATTTACTTGC 2340
QY 2341 GTTGTAAAAGGAAGTAATCCTTTCATCTGTAAGTAAAGGAAAAATAAAC 2389

Db 2341 GTTGTAAAAGGAAGTAATCCTTTCATCTGTAAGTAAAGGAAAAATAAAC 2389
RESULT 4
ABX06886
ID ABX06886 standard; DNA; 2406 BP.
XX AC ABX06886;
XX 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX S. pneumoniae type 4 strain coding region #1174.
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX Streptococcus pneumoniae; type 4 strain.
XX W0200277021-A2.
XX 03-OCT-2002.
XX 27-MAR-2002; 2002WO-IB002163.
XX 27-MAR-2001; 2001GB-00007658.
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX Masignani V, Tettelin H, Fraser C;
XX WPI; 2003-040579/03.
DR P-PSDB; ABU01598.
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX ear infection.
XX Claim 6; SEQ ID NO 2347; 56pp; English.
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2499 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB56454. Also included are an antibody which binds one of the proteins,
XX treating a patient by administering the protein, DNA or antibody (in a
XX composition), a kit comprising first and second primers, which are the
XX nucleic acid cited above or fragments between nucleotides 8-100 of a
XX sequence not defined in the specification, for amplifying a target
XX the first primer is substantially complementary to the target sequence
XX and the second primer is substantially complementary to the complement of
XX the target sequence, and where the parts of the primers having
XX substantial complementarity define the termini of the target sequence to
XX be amplified, assay comprising contacting a test compound with the
XX protein, and determining whether the test compound binds to the protein
XX and a Streptococcus pneumoniae bacterium, where one or more genes
XX encoding the proteins has been rendered inactive. The proteins, nucleic
XX acid molecules, antibody and compositions are useful as medicaments for
XX treating or preventing a disease or infection due to streptococcus
XX bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
XX media or ear infection. They are also useful in developing vaccines,
XX diagnostics and antibiotics. The methods are useful for identifying
XX immunodominant proteins. The present sequence is one of the 2489
XX identified coding region from the genomic sequence. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to

[illegible]

Db 2058 TGGATGGGCAATGCCAGTCAGCATGTGTTAGSCAAGAAAGACCACAGTCAGATCCAAA 2117
 QY 2101 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
 Db 2118 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2177
 QY 2161 TCAAGTAGAGACTCAAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTGTGTCGAA 2220
 Db 2178 TCAAGTAGAGACTCAAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTGTGTCGAA 2237
 QY 2221 AGTAACGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 Db 2238 AGTAACGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2297
 QY 2281 TTTGACTCTTCAATATGATGATCAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2340
 Db 2298 TTTGACTCTTCAATATGATGATCAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2357
 QY 2341 GTTGTAAAAGGAGTAACTCTTCACTGTAAGTAAAGGAAAAATAAAC 2389
 Db 2358 GTTGTAAAAGGAGTAACTCTTCACTGTAAGTAAAGGAAAAATAAAC 2406

RESULT 5

AAA47604
 ID AAA47604 standard; DNA; 2451 BP.
 XX AC AAA47604;
 XX 20-OCT-2000 (first entry)
 XX Recombinant variant of Sp36 gene (Sp36A) of *S. pneumoniae*.
 XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 XX histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX Streptococcus pneumoniae.

Key Location/Qualifiers
 CDS 1..2451
 FT /*tag= a
 FT /product= "Sp36A polypeptide"

XX WO2000037105-A2.
 XX 29-JUN-2000.
 XX 21-DEC-1999; 99WO-US030390.
 XX 21-DEC-1998; 98US-0113048P.
 XX (MEDI-) MEDIMUNE INC.
 XX Johnson LS, Koenig S, Adamou JE;
 XX WPI; 2000-452129/39.
 DR P-PSDB; AAB01468.
 XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX Disclosure; Page 64-65; 70pp; English.

XX Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of *Streptococcus pneumoniae*, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of *S. pneumoniae* in order to utilize such polypeptides in vaccines to
 CC protect against a wide variety of *S. pneumoniae*. New vaccine compositions
 CC are described which comprise a *Streptococcus pneumoniae* polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one

CC histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by *Streptococcus pneumoniae*. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 0 U; 1 Other;
 Query Match 100.0%; Score 2388; DB 3; Length 2451;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 60
 Db 60 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTA 119
 QY 61 TATAGATGGAACACAGCGACGCAAAAACGGAGATTTCGACTCCTCATGAGGTTAGCAA 120
 Db 120 TATAGATGGAACACAGCGACGCAAAAACGGAGATTTCGACTCCTCATGAGGTTAGCAA 179
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGTCAC 180
 Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGTCAC 239
 QY 181 TTCATGCGGACCACTATCATTTATCAATGGTAGGTTCTTATGACGCTATCATCAG 240
 Db 240 TTCATGCGGACCACTATCATTTATCAATGGTAGGTTCTTATGACGCTATCATCAG 299
 QY 241 TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGATATGTTAATGA 300
 Db 300 TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGATATGTTAATGA 359
 QY 301 GGTCAAGGTTGGATATGTTATCAAGTAGTAGTGAATACTATGTTACCTTAAGGATGC 360
 Db 360 GGTCAAGGTTGGATATGTTATCAAGTAGTAGTGAATACTATGTTACCTTAAGGATGC 419
 QY 361 TGCCACGCGGATAAOCGTCCGTACAAAAGAGGAATCAATCGACAAAAACAGAGCATAG 420
 Db 420 TGCCACGCGGATAAOCGTCCGTACAAAAGAGGAATCAATCGACAAAAACAGAGCATAG 479
 QY 421 TCACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTCTGCTTGGCACCCTTCCCA 480
 Db 480 TCACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTCTGCTTGGCACCCTTCCCA 539
 QY 481 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
 Db 540 AGGACGCTATACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 599
 QY 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATACCATTTACATTCTTAAGATGAGTT 600
 Db 600 TGGTGATGCTTATATCGTTCCTCATGGAGATCATACCATTTACATTCTTAAGATGAGTT 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTCTATCTGCTGAGGAAATCTCTCAA 660
 Db 660 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTCTATCTGCTGAGGAAATCTCTCAA 719
 QY 661 TTCAGAAACCTATCGCCGACAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC 720
 Db 720 TTCAGAAACCTATCGCCGACAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC 779
 QY 721 TGTAAAGCAATCCAGAACTACAAATACATAACAGCAACACAGCAACACTTAACAGTCA 780
 Db 780 TGTAAAGCAATCCAGAACTACAAATACATAACAGCAACACAGCAACACTTAACAGTCA 839
 QY 781 AGCAAGTCAAAGTAATGACATGTAGTCTCTTGAACAGCTCTTCAAACTGCCCTTTGAG 840
 Db 840 AGCAAGTCAAAGTAATGACATGTAGTCTCTTGAACAGCTCTTCAAACTGCCCTTTGAG 899
 QY 841 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAACAGTCCGAC 900
 Db 900 TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAACAGTCCGAC 959

901 AGCTAGAGGTGTTGACGTGCGACACGAGATCAITACCACTTCATCCCTTACTCTCAAT 960
 960 AGCTAGAGGTGTTGACGTGCGACACGAGATCAITACCACTTCATCCCTTACTCTCAAT 1019
 961 GTCTGAATTTGGAAGAAAGAAATCGCTCGTATATATTCCTCCTTCTGTTATCGTTTCAAAACCATTTG 1020
 1020 GTCTGAATTTGGAAGAAAGAAATCGCTCGTATATATTCCTCCTTCTGTTATCGTTTCAAAACCATTTG 1079
 1021 GGTACACAGATTCAAGGCGACAAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1080
 1080 GGTACACAGATTCAAGGCGACAAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGG 1139
 1081 CCGCAACCTGCGACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT 1140
 1140 CCGCAACCTGCGACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGT 1199
 1141 ACGAAAGTTGGGGAAGGATATGTTTGAAGAAAGGGGATCTCTCGTTATGTTCTTTC 1200
 1200 ACGAAAGTTGGGGAAGGATATGTTTGAAGAAAGGGGATCTCTCGTTATGTTCTTTC 1259
 1201 GAAAGATTTTACCATTGAAACTGTTTAAATCTTTGAAAGCAAGTTATCAAAACCAAGAGAG 1260
 1260 GAAAGATTTTACCATTGAAACTGTTTAAATCTTTGAAAGCAAGTTATCAAAACCAAGAGAG 1319
 1261 TGTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 1320
 1320 TGTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTTA 1379
 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGGTGC 1380
 1380 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGGTGC 1439
 1381 TAATTCGTATTTCCAAAGCCTTTAGACAAATATTTAGAACGCTTCAATGATGAATCGACTAA 1440
 1440 TAATTCGTATTTCCAAAGCCTTTAGACAAATATTTAGAACGCTTCAATGATGAATCGACTAA 1499
 1441 TAAAGAAATTTGGTATGATGATTTATGGCAATTCCTAGCAACCAATACCCATCCAGAGCG 1500
 1500 TAAAGAAATTTGGTATGATGATTTATGGCAATTCCTAGCAACCAATACCCATCCAGAGCG 1559
 1501 ACTTGCACAAACCAATCTCAATTCAGTATATCTGAAGAGAAAGTTCTGTTATGCTCAAT 1560
 1560 ACTTGCACAAACCAATCTCAATTCAGTATATCTGAAGAGAAAGTTCTGTTATGCTCAAT 1619
 1561 AGCTGATAAGTATACAACTGATGATGATTTTGTGATGAACATGATATATCATGTCG 1620
 1620 AGCTGATAAGTATACAACTGATGATGATTTTGTGATGAACATGATATATCATGTCG 1679
 1621 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAGAGATAG 1680
 1680 TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAGAGATAG 1739
 1681 CTTTCTGATAAGGAAAGTTGCGAGCTCAAGCCTATATCAAGAAAGAAAGGTTATCTTACC 1740
 1740 CTTTCTGATAAGGAAAGTTGCGAGCTCAAGCCTATATCAAGAAAGAAAGGTTATCTTACC 1799
 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGAGCTATTTTA 1800
 1800 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGAGCTATTTTA 1859
 1801 CAATCGTGAAGGGAAGAAAGCAATTCCTGCTGAGCTTCCATATATGTTGAGCA 1860
 1860 CAATCGTGAAGGGAAGAAAGCAATTCCTGCTGAGCTTCCATATATGTTGAGCA 1919
 1861 TACAGTTGAGGTTTAAACCGGTAATTTGATTTTCTCATAGGATCATACCATATAT 1920
 1920 TACAGTTGAGGTTTAAACCGGTAATTTGATTTTCTCATAGGATCATACCATATAT 1979
 1921 TAAATTTGCTGTTTGTATGATCAACATACCAAGCTCCAAATGCTATACCTTCGAAGA 1980
 1980 TAAATTTGCTGTTTGTATGATCAACATACCAAGCTCCAAATGCTATACCTTCGAAGA 2039
 1981 TTTGTTGCGAGGATTAAGTACTAGTAGAACACCCCTGACGAACGTTCCACATTTCTTAATGA 2040

Db 2040 TTTGTTTGGACGATTAAGTACTAGTAGAACACCCCTGACGAAGTCCACATTTCTTAATGA 2099
 Qy 2041 TGGATGGGCAATGCCAGTGAAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA 2100
 Db 2100 TGGATGGGCAATGCCAGTGAAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAA 2159
 Qy 2101 TAAAGACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAAGTCCC 2160
 Db 2160 TAAAGACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAAGTCCC 2219
 Qy 2161 TCAAGTAGAGACTCAAAAGCTAGAGCCCAACTCAAGAAAGACAGAGTTTGTCTGGAA 2220
 Db 2220 TCAAGTAGAGACTCAAAAGCTAGAGCCCAACTCAAGAAAGACAGAGTTTGTCTGGAA 2279
 Qy 2221 AGTAACGGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
 Db 2280 AGTAACGGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2339
 Qy 2281 TTTGACTCTTCAAATATATGATGAATCAATAGTATCATGGCAGAGCAAGAAATTTACTTGC 2340
 Db 2340 TTTGACTCTTCAAATATATGATGAATCAATAGTATCATGGCAGAGCAAGAAATTTACTTGC 2399
 Qy 2341 GTTGTAAAAGGAGTAATCTCTCATCTGTAAGTAAAGAAATAAATAAAC 2389
 Db 2400 GTTGTAAAAGGAGTAATCTCTCATCTGTAAGTAAAGAAATAAATAAAC 2448

RESULT 6

AAV52227
 ID AAV52227 standard; DNA; 8195 BP.
 XX
 AC AAV52227;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US019588.
 XX
 PR 31-OCT-1996; 96US-0029960P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Kunsch CA, Choi GH, Dillion PJ, Rosen CA, Barash SC, Fannon M;
 PI Dougherty BA;
 XX
 DR WPI; 1998-272225/24.
 XX

Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
 Claim 1; Page 727-732; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a

CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC	391, identifying members of the library which contain sequences that
CC	hybridize to the target sequence and isolating the nucleic acid molecules
CC	from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC	organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC	homologous to amplification primers derived from the fragment of the S.
CC	pneumoniae genome to prime the amplification and isolating the amplified
CC	sequences. The computer readable medium can be used in a computer-based
CC	system for identifying fragments of the S. pneumoniae genome of
CC	commercial importance, or expression modulating fragments of the S.
CC	pneumoniae genome. Products from the present invention can be used in
CC	diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC	for S. pneumoniae
XX	
SQ	Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;
	Query Match 100.0%; Score 2388; DB 2; Length 8195;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2388; Conservative
	Mismatches 1; Indels 0; Gaps 0;
QY	1 TTCTTACGAGTTGGACTGTATCAAGCTPAGAACGGTTAAGGAAAATAATCGTGTTCTTA 60
DB	3053 TTCTTACGAGTTGGACTGTATCAAGCTPAGAACGGTTAAGGAAAATAATCGTGTTCTTA 3112
QY	61 TATTAGATGGAATAACAAGCAGCGCAAAAAACGGAGAAATTTGTCCTCATGAGGTTAGCAA 120
DB	3113 TATTAGATGGAATAACAAGCAGCGCAAAAAACGGAGAAATTTGTCCTCATGAGGTTAGCAA 3172
QY	121 GCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGTATACAGCAAGGCTATGTCAC 180
DB	3173 GCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGTATACAGCAAGGCTATGTCAC 3232
QY	181 TTCACATGCGCACCACTATCATTTATCAATGGTAAGGTTCTTATGACGCTATCATCAG 240
DB	3233 TTCACATGCGCACCACTATCATTTATCAATGGTAAGGTTCTTATGACGCTATCATCAG 3292
QY	241 TGAGAAATTTACTCATGAAAGATCCAAAATATAAGCTAAAAGTAGAGATATTGTTTAATGA 300
DB	3293 TGAGAAATTTACTCATGAAAGATCCAAAATATAAGCTAAAAGTAGAGATATTGTTTAATGA 3352
QY	301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACATGTTTACCCTTAAGGATGC 360
DB	3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAATACATGTTTACCCTTAAGGATGC 3412
QY	361 TGCCCCACGGGATAACGTCGGTCACAAAAGAGAGAAATCAATCGACAAAAAACAGACATAG 420
DB	3413 TGCCCCACGGGATTAACGTCGGTCACAAAAGAGAGAAATCAATCGACAAAAAACAGACATAG 3472
QY	421 TCACATCTGTAAGGTGGAACTCCAAGAAACGATGGTGCTGTGTTCCTTGGCACGTTCCGA 480
DB	3473 TCACATCTGTAAGGTGGAACTCCAAGAAACGATGGTGCTGTGTTCCTTGGCACGTTCCGA 3532
QY	481 AGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB	3533 AGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592
QY	541 TGGTGATGCTTATATCGTTCCCTCATGGAGATCATTTACCATTACATTCCTTAAGATGAGTT 600
DB	3593 TGGTGATGCTTATATCGTTCCCTCATGGAGATCATTTACCATTACATTCCTTAAGATGAGTT 3652
QY	601 ATCAGCTAGCGATTGGCTGCTGCAAGAGCCTTCCTATCTGTGCGAGGAAATCTGTCAAA 660
DB	3653 ATCAGCTAGCGATTGGCTGCTGCAAGAGCCTTCCTATCTGTGCGAGGAAATCTGTCAAA 3712
QY	661 TTCAAGAACCTATCGGCCGACAAATAGCGNATACATTCACAAGCAAACTGGGTACCTTC 720
DB	3713 TTCAAGAACCTATCGGCCGACAAATAGCGNATACATTCACAAGCAAACTGGGTACCTTC 3772
QY	721 TGTAAGCAATCCAGGAACTACAATACTATCAACAAGCAACAAACAGCAACACTAACAGTCA 780
DB	3773 TGTAAGCAATCCAGGAACTACAATACTATCAACAAGCAACAAACAGCAACACTAACAGTCA 3832
QY	781 AGCAAGTCAAAGTAATGACATTTGANTAGTCTCTTTGAAACAGCTCTACAAACTGCCTTTGAG 840

Db	4913	TACAGTTGAGGTTAAACCGTAAATTTGATTATTCCTCATAGGATCATACCAATAAT	4972
Qy	1921	TAAATTTGCTTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATATCTTGGAGA	1980
Db	4973	TAAATTTGCTTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATATCTTGGAGA	5032
Qy	1981	TTTGTGTCGACGATTAAAGTACTAGTAGAACAACCTGACGAAAGTCCACACATTCCTAATGA	2040
Db	5033	TTTGTGTCGACGATTAAAGTACTAGTAGAACAACCTGACGAAAGTCCACACATTCCTAATGA	5092
Qy	2041	TGGATGGGCAATCCGATGAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAA	2100
Db	5093	TGGATGGGCAATCCGATGAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAA	5152
Qy	2101	TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGTCGAGCCAGAGTCCC	2160
Db	5153	TAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGTCGAGCCAGAGTCCC	5212
Qy	2161	TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTCTTGGAA	2220
Db	5213	TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAAAGTTTGTCTTGGAA	5272
Qy	2221	AGTAACCGATTCTAGTCTGAAGCCCAATGCACAGAACTCTAGCTGGTTTACGAAATAA	2280
Db	5273	AGTAACCGATTCTAGTCTGAAGCCCAATGCACAGAACTCTAGCTGGTTTACGAAATAA	5332
Qy	2281	TTTGACTCTTCAAATATTGGATAACAATAGTAGTATCATGGCAGAGCAAGAAATTAAC	2340
Db	5333	TTTGACTCTTCAAATATTGGATAACAATAGTAGTATCATGGCAGAGCAAGAAATTAAC	5392
Qy	2341	GTTGTTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGTAAAGGAAATAAAC	2389
Db	5393	GTTGTTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGTAAAGGAAATAAAC	5441

RESULT 7

ABSS6454_09/c
Continuation (10 of 22) of ABSS6454 from base 900001 (Streptococcus pneumoniae type 4 sd
WP Sequence split into 22 fragments LOCUS ABSS6454 Accession ABSS6454

WP	Fragment Name	Begin	End
WP	ABSS6454_01	100001	110000
WP	ABSS6454_02	200001	210000
WP	ABSS6454_03	300001	310000
WP	ABSS6454_04	400001	410000
WP	ABSS6454_05	500001	510000
WP	ABSS6454_06	600001	610000
WP	ABSS6454_07	700001	710000
WP	ABSS6454_08	800001	810000
WP	ABSS6454_09	900001	910000
WP	ABSS6454_10	1000001	1010000
WP	ABSS6454_11	1100001	1110000
WP	ABSS6454_12	1200001	1210000
WP	ABSS6454_13	1300001	1310000
WP	ABSS6454_14	1400001	1410000
WP	ABSS6454_15	1500001	1510000
WP	ABSS6454_16	1600001	1610000
WP	ABSS6454_17	1700001	1710000
WP	ABSS6454_18	1800001	1810000
WP	ABSS6454_19	1900001	1910000
WP	ABSS6454_20	2000001	2010000
WP	ABSS6454_21	2100001	2110000
WP	ABSS6454_21	2100001	2162598

Query Match
Best Local Similarity 100.0%; Score 2388; DB 7; Length 110000;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	TTCTACGATTTGGGACTGTATCAAGCTAGAGCGTTAAGGAAATAATCGTGTTCCTA	60
Db	106950	TTCTACGATTTGGGACTGTATCAAGCTAGAGCGTTAAGGAAATAATCGTGTTCCTA <td>106891</td>	106891
Qy	61	TATAGATGAAACACAGCCACCAAAACCGAGAAATTTGATCTCTGATAGGTTAGCAA	120

Db	106890	TATAGATGAAACACAGCCACCAAAACCGAGAAATTTGATCTCTGATAGGTTAGCAA	106831
Qy	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAAAGACCAAGGCTATGTCCAC	180
Db	106830	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAAAGACCAAGGCTATGTCCAC	106771
Qy	181	TTTCATGCGGACCACTATCATTTAATTAATGGAAGGTTCTTATGACGCTATCATCAG	240
Db	106770	TTTCATGCGGACCACTATCATTTAATTAATGGAAGGTTCTTATGACGCTATCATCAG	106711
Qy	241	TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA	300
Db	106710	TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA	106651
Qy	301	GGTCAAGGGTGGATGATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Db	106650	GGTCAAGGGTGGATGATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	106591
Qy	361	TGCCACGCGGATACCGTCCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
Db	106590	TGCCACGCGGATACCGTCCGTACAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	106531
Qy	421	TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGCTGCTGCTTGGCCTTGGCAGTTCGCA	480
Db	106530	TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGCTGCTGCTTGGCCTTGGCAGTTCGCA	106471
Qy	481	AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGGATATC	540
Db	106470	AGGACGCTATCTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGGATATC	106411
Qy	541	TGCTGATGCTTATATGTTTCTCATGAGATCAATACCATTAATTCCTAAGAAATGATTT	600
Db	106410	TGCTGATGCTTATATGTTTCTCATGAGATCAATACCATTAATTCCTAAGAAATGATTT	106351
Qy	601	ATCAGCTAGCGAGTGTGCTGCTGAGAAAGCTTCTTATCTGCTGCGAGGAATCTGTCAA	660
Db	106350	ATCAGCTAGCGAGTGTGCTGCTGAGAAAGCTTCTTATCTGCTGCGAGGAATCTGTCAA	106291
Qy	661	TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAAACAACTGCGGTACTTC	720
Db	106290	TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAAACAACTGCGGTACTTC	106231
Qy	721	TGTAAGCAATCCAGGAACTACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	780
Db	106230	TGTAAGCAATCCAGGAACTACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	106171
Qy	781	AGCAAGTCAAAGTAATGATGATGCTCTTGAACAGCTCTACAACTGCTGCTTGG	840
Db	106170	AGCAAGTCAAAGTAATGATGATGCTCTTGAACAGCTCTACAACTGCTGCTTGG	106111
Qy	841	TCAAAGCAATGTAGAAATCTGATGGGCTTCTTTGATCCAGCAACAAATCACAAGTCCAA	900
Db	106110	TCAAAGCAATGTAGAAATCTGATGGGCTTCTTTGATCCAGCAACAAATCACAAGTCCAA	106051
Qy	901	AGCTAGAGGTGTTGAGTCCGACAGGATCAATACCACTTCCCTTACTCTCAAT	960
Db	106050	AGCTAGAGGTGTTGAGTCCGACAGGATCAATACCACTTCCCTTACTCTCAAT	105991
Qy	961	GTCTGAATTTGGAAGAACGAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATTTG	1020
Db	105990	GTCTGAATTTGGAAGAACGAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATTTG	105931
Qy	1021	GGTACAGATTCAGCCGAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG	1080
Db	105930	GGTACAGATTCAGCCGAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG	105871
Qy	1081	CCCGCAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTGTAGTCACTGCT	1140
Db	105870	CCCGCAACTCGACCAAAATCTTAAATAGACTCAAAATCTCTTTGGTGTAGTCACTGCT	105811
Qy	1141	ACGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGCTTTGC	1200
Db	105810	ACGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGCTTTGC	105751

QY	1201	GAAAGATTACCACTGTAAGAACTGTTAAATAATCTTGAAAGCAAGTATCAAAAACAGAGAG	1260
Db	105750	GAAAGATTACCACTGTAAGAACTGTTAAATAATCTTGAAAGCAAGTATCAAAAACAGAGAG	105691
QY	1261	TGTTTACACACTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1320
Db	105690	TGTTTACACACTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA	105631
QY	1321	TGATAAAGCAATATAATCTGTTAACTGAGGCTCATATAAGCCCTTGTTGNAATAAAGGCTCG	1380
Db	105630	TGATAAAGCAATATAATCTGTTAACTGAGGCTCATATAAGCCCTTGTTGNAATAAAGGCTCG	105571
QY	1381	TAATCTTGATTTCAAGCCTTAGACAAATTAATTAAGAACGCTTGATGATGATGATGATGAT	1440
Db	105570	TAATCTTGATTTCAAGCCTTAGACAAATTAATTAAGAACGCTTGATGATGATGATGATGAT	105511
QY	1441	TAAAGAAATTTGCTAGATGATTTATTGGCAATCTAGACCAATATACCCATCCAGAGCG	1500
Db	105510	TAAAGAAATTTGCTAGATGATTTATTGGCAATCTAGACCAATATACCCATCCAGAGCG	105451
QY	1501	ACTTGGCAAAACCAATTTCTCAAAATTTGATTAATCTGAAGACGAAGTTCGTATTGCTCAATT	1560
Db	105450	ACTTGGCAAAACCAATTTCTCAAAATTTGATTAATCTGAAGACGAAGTTCGTATTGCTCAATT	105391
QY	1561	AGCTGATAAGATATACAGCTGAGATGTTTATTTGATGAACATGATATAATCAGTGA	1620
Db	105390	AGCTGATAAGATATACAGCTGAGATGTTTATTTGATGAACATGATATAATCAGTGA	105331
QY	1621	TGAAGGATGATATGTAAGTAAAGCCCTCATATGGCCATAGTCACTGGATGGAAAGATAG	1680
Db	105330	TGAAGGATGATATGTAAGTAAAGCCCTCATATGGCCATAGTCACTGGATGGAAAGATAG	105271
QY	1681	CTTCTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTTATCTAAGAAAAAGGTATCCTACC	1740
Db	105270	CTTCTTCTGATTAAGGAAAAAGTTGACGCTCAAGCCTTATCTAAGAAAAAGGTATCCTACC	105211
QY	1741	TCCATCTCCAGCGGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGAGCTATTTA	1800
Db	105210	TCCATCTCCAGCGGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGAGCTATTTA	105151
QY	1801	CAATCGTGTGAAGGGGAAAAACGAATCCACTGCTTGACCTTCCATATATGTTGGAGCA	1860
Db	105150	CAATCGTGTGAAGGGGAAAAACGAATCCACTGCTTGACCTTCCATATATGTTGGAGCA	105091
QY	1861	TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTAAGGATCATTTACCAATAT	1920
Db	105090	TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTAAGGATCATTTACCAATAT	105031
QY	1921	TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1980
Db	105030	TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	104971
QY	1981	TTTGTGTCGACGATTAAGTACTACGTAGAACACCCCTGACGACGCTCCATTTCTAATGA	2040
Db	104970	TTTGTGTCGACGATTAAGTACTACGTAGAACACCCCTGACGACGCTCCATTTCTAATGA	104911
QY	2041	TGATGGGGCAATGTCAGGTGAGCATGTTTAGCAAGAAGACCAAGTGAAGATCCAAA	2100
Db	104910	TGATGGGGCAATGTCAGGTGAGCATGTTTAGCAAGAAGACCAAGTGAAGATCCAAA	104851
QY	2101	TAAAGACTTCAAGCGGATGAAGCGCAGTGAAGGAAACACCTGCTGAGCCGAAGTCCC	2160
Db	104850	TAAAGACTTCAAGCGGATGAAGCGCAGTGAAGGAAACACCTGCTGAGCCGAAGTCCC	104791
QY	2161	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGATTTTGTTCGGAA	2220
Db	104790	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGATTTTGTTCGGAA	104731
QY	2221	AGTAAGCGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA	2280
Db	104730	AGTAAGCGGATTTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATAA	104671

QY	2281	TTTGACTCTTCAAAATTTATGATAAACAATAGTATCATGCGCAAGACAGAAAAATTTACTTGC	2340
Db	104670	TTTGACTCTTCAAAATTTATGATAAACAATAGTATCATGCGCAAGACAGAAAAATTTACTTGC	104611
QY	2341	GTTGTTAAAGGAAGTAATCTTCACTCTGAAGTAAGGAAAAATAAAC	2389
Db	104610	GTTGTTAAAGGAAGTAATCTTCACTCTGAAGTAAGGAAAAATAAAC	104562

RESULT 8
 ABS56454 10/c
 Continuation (11 of 22) of ABS56454 from base 1000001 (Streptococcus pneumoniae type 4 st
 WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454
 Fragment Name Begin End
 WP ABS56454_00 1 110000
 WP ABS56454_01 100001 210000
 WP ABS56454_02 200001 310000
 WP ABS56454_03 300001 410000
 WP ABS56454_04 400001 510000
 WP ABS56454_05 500001 610000
 WP ABS56454_06 600001 710000
 WP ABS56454_07 700001 810000
 WP ABS56454_08 800001 910000
 WP ABS56454_09 900001 1010000
 WP ABS56454_10 1000001 1110000
 WP ABS56454_11 1100001 1210000
 WP ABS56454_12 1200001 1310000
 WP ABS56454_13 1300001 1410000
 WP ABS56454_14 1400001 1510000
 WP ABS56454_15 1500001 1610000
 WP ABS56454_16 1600001 1710000
 WP ABS56454_17 1700001 1810000
 WP ABS56454_18 1800001 1910000
 WP ABS56454_19 1900001 2010000
 WP ABS56454_20 2000001 2110000
 WP ABS56454_21 2100001 2162598

Query Match 100.0%; Score 2388; DB 7; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTCTTACAGTTGGAGCTGTATCAAGCTAGAACCGTAAAGAAAAATAATCGTGTTCCTTA	60
Db	6950	TTCTTACAGTTGGAGCTGTATCAAGCTAGAACCGTAAAGAAAAATAATCGTGTTCCTTA	6891
QY	61	TATAGTGAAGAAACAAAGCGACGAAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	120
Db	6890	TATAGTGAAGAAACAAAGCGACGAAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	6831
QY	121	GGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCGGCTATCTCAC	180
Db	6830	GGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCGGCTATCTCAC	6771
QY	181	TTTCATGCGGACCACTCATTTATCAATGTTAGGTTTCTTATGACGCTATCATCAG	240
Db	6770	TTTCATGCGGACCACTCATTTATCAATGTTAGGTTTCTTATGACGCTATCATCAG	6711
QY	241	TGAAGAAATTTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAATGA	300
Db	6710	TGAAGAAATTTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAATGA	6651
QY	301	GGTCAAGGTTGATATTATCAAGTAGATGGAATAACTATGTTTACTCTTAAGGATGC	360
Db	6650	GGTCAAGGTTGATATTATCAAGTAGATGGAATAACTATGTTTACTCTTAAGGATGC	6591
QY	361	TGCCCCACGGGATAAGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG	420
Db	6590	TGCCCCACGGGATAAGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG	6531
QY	421	TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTTGGCTTGGCCTTGGCAGCTTCGCA	480
Db	6530	TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTTGGCTTGGCCTTGGCAGCTTCGCA	6471

481 AGGACGCTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGTAC 540
|||||
6470 AGGACGCTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGTAC 6411
|||||
541 TGGTATGCTTATATCGTTCTCATGAGATCATATACCAATTAATTCCTAAGAAATGAGTT 600
|||||
6410 TGGTATGCTTATATCGTTCTCATGAGATCATATACCAATTAATTCCTAAGAAATGAGTT 6351
|||||
601 ATCAGTACGAGTTGGCTGCTCAGAGCTTCCTATCTGTTGAGGAAATCTGTCAAA 660
|||||
6350 ATCAGTACGAGTTGGCTGCTCAGAGCTTCCTATCTGTTGAGGAAATCTGTCAAA 6291
|||||
661 TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
|||||
6290 TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 6231
|||||
721 TGTAAAGCAATCCAGGAATCAAAATACTAACCAAGCAACAAAGCAAACTAACAGTCA 780
|||||
6230 TGTAAAGCAATCCAGGAATCAAAATACTAACCAAGCAACAAAGCAAACTAACAGTCA 6171
|||||
781 AGCAAGTCAAGTAAATGACATTAATGATGCTCTTGAAGACAGCTTCAAACTGCTTTGAG 840
|||||
6170 AGCAAGTCAAGTAAATGACATTAATGATGCTCTTGAAGACAGCTTCAAACTGCTTTGAG 6111
|||||
841 TCAACGACATGTAGAACTCTGATGGCCTTCTCTTGTATCCAGCACAAATCACAAAGTCGAAC 900
|||||
6110 TCAACGACATGTAGAACTCTGATGGCCTTCTCTTGTATCCAGCACAAATCACAAAGTCGAAC 6051
|||||
901 AGCTAGAGTGTGAGTGCACAGGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 960
|||||
6050 AGCTAGAGTGTGAGTGCACAGGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 5991
|||||
961 GTCTCAATTTGGAAGACGAATCGCTGTAATTAATCCCTTCCTGTTATCTGTTCAACCATTTG 1020
|||||
5990 GTCTCAATTTGGAAGACGAATCGCTGTAATTAATCCCTTCCTGTTATCTGTTCAACCATTTG 5931
|||||
1021 GGTACCAGATTCAGGCCAGAAACAAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAGG 1080
|||||
5930 GGTACCAGATTCAGGCCAGAAACAAACCAAGTCCACAAACGACTCCGGAACCTAGTCCAGG 5871
|||||
1081 CCGGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGTT 1140
|||||
5870 CCGGCAACTGCAACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGTT 5811
|||||
1141 AGCAAGTGTGGGAGGATGATGTTTGAAGAAAGGCACTCTCGTTATGCTTTTGC 1200
|||||
5810 AGCAAGTGTGGGAGGATGATGTTTGAAGAAAGGCACTCTCGTTATGCTTTTGC 5751
|||||
1201 GAAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 1260
|||||
5750 GAAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 5691
|||||
1261 TGTTCACACATTTAATCTGTTTAAATAGAAATGTTGCTCTCGTGAACCAAGAAATTTTA 1320
|||||
5690 TGTTCACACATTTAATCTGTTTAAATAGAAATGTTGCTCTCGTGAACCAAGAAATTTTA 5631
|||||
1321 TGATAAGCATATAATCTGTTTAACTGAGCTCATTAAGCCTGTTTGNAAATTAAGGTCG 1380
|||||
5630 TGATAAGCATATAATCTGTTTAACTGAGCTCATTAAGCCTGTTTGNAAATTAAGGTCG 5571
|||||
1381 TAAATCTGATTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATCACTAA 1440
|||||
5570 TAAATCTGATTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATCACTAA 5511
|||||
1441 TAAAGAAAAATTTGGTAGATTTATTTGGCAATCTTAGACAAATTTACCCATCCAGAGCG 1500
|||||
5510 TAAAGAAAAATTTGGTAGATTTATTTGGCAATCTTAGACAAATTTACCCATCCAGAGCG 5451
|||||
1501 ACTTGGCAACCAATTTCTCAAAATGAGTACTGAAGACGAGTTTCGTATGCTCAATT 1560
|||||
5450 ACTTGGCAACCAATTTCTCAAAATGAGTACTGAAGACGAGTTTCGTATGCTCAATT 5391
|||||
1561 AGCTGATAAGTATACAAGCTCAGATGGTTTACATTTTGTATGAACATGATAATCAGTGA 1620
|||||

Db 5390 AGCTGATAAGTATACAACGCTCAGATGGTTTACATTTTGTATGAACATGATATAATCAGTGA 5331
|||||
QY 1621 TGAAGGAGATGCAATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 1680
|||||
Db 5330 TGAAGGAGATGCAATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAAAAGATAG 5271
|||||
QY 1681 CTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATCTTAAAGAAAAAGGTTATCTTACC 1740
|||||
Db 5270 CTTTCTGATAAGGAAAAAGTTGCGAGCTCAAGCCTATCTTAAAGAAAAAGGTTATCTTACC 5211
|||||
QY 1741 TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
|||||
Db 5210 TCCATCTCCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 5151
|||||
QY 1801 CAATCTGTGAAAGGGAAAAACGAAATCCACTTCGTTCCACTTCCATATATGTTGTCAGCA 1860
|||||
Db 5150 CAATCTGTGAAAGGGAAAAACGAAATCCACTTCGTTCCACTTCCATATATGTTGTCAGCA 5091
|||||
QY 1861 TACAGTTGAGGTTTAAACCGGTAATTTGATTAATCTCTCAATAAGGATCAATTACCATTAAT 1920
|||||
Db 5090 TACAGTTGAGGTTTAAACCGGTAATTTGATTAATCTCTCAATAAGGATCAATTACCATTAAT 5031
|||||
QY 1921 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAGA 1980
|||||
Db 5030 TAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAGA 4971
|||||
QY 1981 TTTGTTTGGCGAGTTAAGTACTAGTGTAGAACACCCCTGACGAACGTCACATTTCTTAATCA 2040
|||||
Db 4970 TTTGTTTGGCGAGTTAAGTACTAGTGTAGAACACCCCTGACGAACGTCACATTTCTTAATCA 4911
|||||
QY 2041 TGGATGGGGCAATGCGAGTGAATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
|||||
Db 4910 TGGATGGGGCAATGCGAGTGAATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 4851
|||||
QY 2101 TAAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGCAACACCTGCTGAGCCAGAGTCCC 2160
|||||
Db 4850 TAAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGCAACACCTGCTGAGCCAGAGTCCC 4791
|||||
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAACGAGAAATTTTCTTGGCGAA 2220
|||||
Db 4790 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAACGAGAAATTTTCTTGGCGAA 4731
|||||
QY 2221 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAAAATCTAGCTGGTTTACGAATAA 2280
|||||
Db 4730 AGTAAACGATTTAGTCTGAAAGCCAAATGCAACAGAAAATCTAGCTGGTTTACGAATAA 4671
|||||
QY 2281 TTTGACTCTTCAAAATTTATGGATTAACAATAGTATCATGGCAGAACGAGAAATTTACTTGC 2340
|||||
Db 4670 TTTGACTCTTCAAAATTTATGGATTAACAATAGTATCATGGCAGAACGAGAAATTTACTTGC 4611
|||||
QY 2341 GTTGTAAAAAGGAAGTAAATCTTCACTGTAAAGTAAAGAAAAATAAAC 2389
|||||
Db 4610 GTTGTAAAAAGGAAGTAAATCTTCACTGTAAAGTAAAGAAAAATAAAC 4562
|||||

RESULT 9

AAA65731

ID AAA65731 standard; DNA; 2523 Bp.

XX AC AAA65731;

XX XX

DT 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.

DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal; ds.

XX Streptococcus pneumoniae.

XX OS

XX PN

WO200039299-A2.

XX 06-JUL-2000.
 PD 20-DEC-1999; 99MO-CA001218.
 XX 23-DEC-1998; 98US-0113800P.
 XX (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 DR P-PSDB; AAB12716.
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia.
 XX Example 2; Fig 3; 106pp; English.
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
 CC antigen
 XX
 XX Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;
 Query Match 57.5%; Score 1374.2; DB 3; Length 2523;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 QY 1 TTCTTAGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60
 DB 60 TGCTTATGAACAGTGGTTGGTCAATCAAGCTCAAACTGTAAAGAAAATAATCGTGTTCCTA 119
 QY 61 TATAGATGGAAGAACAGGAGCGAAGAACGGAGAAATTTGACTCCTGATGAGGTAGCAA 120
 DB 120 TATAGATGGAAGAACAGGAGCGAAGAACGGAGAAATTTGACTCCTGATGAGGTAGCAA 179
 QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATACAGACCAAGGCTATGTCAC 180
 DB 180 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATACAGATCAAGTTATGTGAC 239
 QY 181 TTCAATGGGACCACTATCATATTATTAATGATGATGATGATGATGATGATGATGATGATG 240
 DB 240 CTCTCATGGAGACATTATCATTTACTATATATATGATGATGATGATGATGATGATGATGATG 299
 QY 241 TGAAGAAATTACTCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA 300
 DB 300 TGAAGAGCTCCTCATGAAGATCCGAATATATGATGATGATGATGATGATGATGATGATGATG 359
 QY 301 GGTCAAGGTCGATATGTTATCAAGGTAGATGGAATACTATGTTACCTTAAGGATGC 360
 DB 360 AATCAAGGTCGATATGTTATCAAGGTAGATGGAATACTATGTTACCTTAAGGATGC 419
 QY 361 TGCCCAACCGGATTAACGTCCTGATCAAAAGAGAAATCAATCAAGAAACAAAGAGCATAG 420
 DB 420 AGCTCATGCGGATAATGTCCTGATCAAAAGAGAAATCAATCAAGAAACAAAGAGCATAG 479
 QY 421 TCACATCGTGAAGTGAAGTCAAGAAACGATGGTGTGCTGTGCTGGCAGCGTTCGCA 480
 DB 480 TCAGCATCGTGAAGGAGGATTCAGCAACATGTCGGTAGGCTTTGACAGCTTCACA 539
 QY 481 AGGACGCTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
 DB 540 GGGACGCTACACCAAGATGATGTTTATATCTTCAATGATCTGATATCATCGAAGATAC 599
 QY 541 TGGTGAATGTTATGTTCTCTCATGAGATCATATACCATTTACATTTCTCTTAAGATGAGTT 600

DB 600 GGCGGATGCTATATCGTTCTCATGGAGATCAATACCATTACATTCCTAAGAAATGAGTT 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTCGAGGAAATCTGTCAA 660
 DB 660 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTCGAGGAAATCTGTCAA 719
 QY 661 TTCAAGAAACCTATCGCCGACAAATAGCGATAACACTCCAAAGAACAACTGGGTACCTTC 720
 DB 720 TTAAAGAAACCTATCGCCGACAAATAGCGATAACACTCCAAAGAACAACTGGGTACCTTC 779
 QY 721 TGTAAAGAACTAGGAACTACAAATATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 780
 DB 780 TGTAAAGCAATCCAGGAACTACAAATATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 839
 QY 781 AGCAAGTCAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTCAAACTGCTTTGAG 840
 DB 840 AGCAAGTCAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTCAAACTGCTTTGAG 899
 QY 841 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATTCAGACAAATCAAAAGTCAAG 900
 DB 900 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATTCAGACAAATCAAAAGTCAAG 959
 QY 901 AGCTAGAGGTGTTGCGAGTGCACACGAGATCAATACCACTTCATCCTTACTCTCAAT 960
 DB 960 CGCAGAGGTGTAGCTGTCCTCATGTAAACCACTTACCACTTTATCCTTTATGACAAAT 1019
 QY 961 GTCTGAATTCGAAAGACGAATCGTCTGATTTATTTCCCTTCGTTATCGTTCAACCAATG 1020
 DB 1020 GTCTGAATTCGAAAGACGAATCGTCTGATTTATTTCCCTTCGTTATCGTTCAACCAATG 1079
 QY 1021 GGTACGAGATTCAGGCGCAAGCAACCAAGTCCCAACCGACTCCGGAACTAGTCCAGG 1080
 DB 1080 GGTACGAGATTCAGGCGCAAGCAACCAAGTCCCAACCGACTCCGGAACTAGTCCAGG 1139
 QY 1081 CCGCACTGCAACCAATCTTAAATAGACTCAA-----ATTCTTCTTGGT 1128
 DB 1140 TCCGCAACCTGCAACCAATCTTAAACAGCTCCAGCAATCCAAATTCATGAGAATTTGGT 1199
 QY 1129 TAGTCAGCTGGTACGAAAGTGGGGAAGGATATGATTCGAAAGAAAGGCACTCTCTCG 1188
 DB 1200 CAAAGAGCTGTTGAAAGTGGGATGTTGTTCTTTGAGGAGAAATGAGTTCTCG 1259
 QY 1189 TTATGCTTTGCGAAGATTTACCATCTGAACTGTTTAAATCTTGAAGCAAGTTATC 1248
 DB 1260 TTATATCCGCAAGAAATCTTTGAGGAGAAACAGCAGCAGGCAATGATAGCAACATGGC 1319
 QY 1249 AAAAAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTA 1308
 DB 1320 CAAAGCAAGATTTATCTCATAGCTAGGAGCTTAAAGAACTGACCTCCCATCTAGTGA 1379
 QY 1309 CCAAGAAATTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTTGTTGN 1368
 DB 1380 TCGAGAAATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGA 1439
 QY 1369 AAATAAGGTCGTAATTTCTGATTTCCAGCTTGAACAATATTATAGACCTTTGATGA 1428
 DB 1440 TAATAAAGGTCGCAAGTTGATTTTGGAGCTTTGGATAAOCCTTGGAAACCACTCAAGGA 1499
 QY 1429 TGAATCGACTAATAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCATAATAC 1488
 DB 1500 TGTCTCAAGTGAATAAGTCAAGTTAGTGAATATTTCTTGGCTTCTTAGCTCCGATTCG 1559
 QY 1489 CCATCCAGAGCCTTGGCAACCAAAATCTCAAAATTTGAGTATATCTGAAGACGAGTTTCG 1548
 DB 1560 TCATCCAGACGTTTAGGAAACCAAAATTCGCAAAATACCTACACTGATGATGATGATGATGAT 1619
 QY 1549 TATTGCTCAATTTAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1608
 DB 1620 AGTACCAAGTTGGCAGGCAAGTACACAAAGAGACGCTTATATCTTTGATTCCTCTGTA 1679
 QY 1609 TATAATCAGTGTGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGAT 1668
 DB 1680 TATAACCAAGTGTGAGGGGATGCTTATGTAATCCACATATGACCCATAGCCTAGCTGAT 1739

QY 1669 TGAAGAGATAGCTTTCTGATAGGAAAAAGTTGCAGCTCAAGCCTATATCTAAAGAAA 1728
 Db |||||
 QY 1740 TAAAAAGATAGTTGTCTGAAGCTGAGAGCGCGAGCCAGGCTTATGCTAAAGAGAA 1799
 Db |||||
 QY 1729 AGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTC 1788
 Db |||||
 QY 1800 AGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACCTGAGGCAAAAGGAGC 1859
 Db |||||
 QY 1789 AGCAGTATTTACATCTGTGTAAGGGGAAAAACGAATTCACCTGCTGAGCTTCCATA 1848
 Db |||||
 QY 1860 AGAAGCTATCTCAACCGCGTGAAGACAGCTAAGAAGGTGCCACITGATCGTATGCTT 1919
 Db |||||
 QY 1849 TATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTCAATAGCATCA 1908
 Db |||||
 QY 1920 CAATCTTCAATATCTGTAGAGTCAAAACGGTAGTTTAACTATACCTCAATATGACCA 1979
 Db |||||
 QY 1909 TTACCATATATTAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTA 1968
 Db |||||
 QY 1980 TTACCATATACATCAAAATTTGAGTGGTTGACGAAGCCCTTTATGAGGCACCTTAAGGGTA 2039
 Db |||||
 QY 1969 TACCTTGGAGATTTGTTTGGACGATTAAGTACTAGCTAGAACCCCTGACGAAGCTCC 2028
 Db |||||
 QY 2040 TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAACGCTC 2099
 Db |||||
 QY 2029 ACATTTCTAATGATGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACAG 2088
 Db |||||
 QY 2100 GCATTCAGATATGTTTGTGTAACGCTAGCNCATGTTCAAGAAACAAATGGTCA 2159
 Db |||||
 QY 2089 TGAAGTCCAAATAGAACTTCAAAGCGGATGAAGAG----- 2125
 Db |||||
 QY 2160 AGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGCA 2219
 Db |||||
 QY 2126 ----- 2125
 Db |||||
 QY 2220 GGAAGAAACCCCTCGAAGAGAGAAACCAAAAGCGAGAAACAGAGTCTCCAAAAACCAAC 2279
 Db |||||
 QY 2126 -----CCAGTAGGAGAAACACCTGCTGAGCGAGAAAGTCCCTCAAGTAGAGACTGAAAA 2178
 Db |||||
 QY 2280 AGAGGAAACCAAGAGAGAGATCACCAGAGGAAATCAGAGAACTCAGTCCAGACTGAAAA 2339
 Db |||||
 QY 2179 AGTAGAGCCCACTCAAAGAGAGAGAGTTTGTGCTGCGAAAGTAAACGGATTTCTAGTCT 2238
 Db |||||
 QY 2340 GGTTCGAAGAAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAATTAT 2399
 Db |||||
 QY 2239 GAAAGCCATGCAACAGAACTCTAGCTGTTTACGAATATTTGACTCTTCAATATAT 2298
 Db |||||
 QY 2400 CAAGTCCCAATGCCAAGAGACTCTCAGAGATTAATAATTTACTATTTTGGCACCCA 2459
 Db |||||
 QY 2299 GGATAACAATAGTATCATGGCAGAGAGCAAAAAATTTACTTGGTGTGTTAAAAAGGAGTAA 2358
 Db |||||
 QY 2460 GGACACAACTATATATGGCAGAGCTGAAAAAATTTATTTGGCTTTTAAAGGAGATTA 2519
 Db |||||
 RESULT 10
 ID AAA65736
 XX AAA65736 standard; DNA; 2647 BP.
 AC
 XX
 XX
 DT 21-NOV-2000 (first entry)
 DE
 DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX
 FN WO200039299-A2.
 XX
 PD 06-JUL-2000.

XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 XX
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 XX
 PS Example 6; Fig 15; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
 CC antigen
 XX
 SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
 Query Match 57.5%; Score 1374.2; DB 3; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
 QY 1 TCTTACGAGTTGGGACTGTATCAAGCTTAGAAGCGTTAAGGAAAAATATCGTGTTCCTA 60
 Db |||||
 QY 104 TCGTTATGAACCTAGGTTTGCATCAAGCTCAAACTGTAAAGAAAAATATCGTGTTCCTA 163
 Db |||||
 QY 61 TATAGATGAAAAAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 120
 Db |||||
 QY 164 TATAGATGAAAAAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 223
 Db |||||
 QY 121 GCGTGAAGAAATCAATGCTGTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCA 180
 Db |||||
 QY 224 GCGTGAAGAAATCAACGCGCAAAATCGTCAATCAAGATTAACGATCAAGGTTATGTGAC 283
 Db |||||
 QY 181 TTCATATGCGGACCACTATCATATTATCAATGTAAGGTTTCTTATGACGCTATCATCAG 240
 Db |||||
 QY 284 CTCTCATGGAGACCAATTATCTACTATAATGCGAAGGTCCTTATGATGCCATCATCAG 343
 Db |||||
 QY 241 TGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAATGA 300
 Db |||||
 QY 344 TGAAGAGCTCTCATGAAAGATCCGAATATCAGTTGAAGGATTCAGACATTTGTCAATGA 403
 Db |||||
 QY 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGAAAAATCTATGTTTACCTTTAAGGATGC 360
 Db |||||
 QY 404 AATCAAGGTTGGTTATGTTCATTAAGGTAAACGGTAAATACTATGTTTACCTTTAAGGATGC 463
 Db |||||
 QY 361 TGCCCAACCGGATACAGTCCGTACAAAAGAGAGAAATCAATCGACAAAAACAAAGACATAG 420
 Db |||||
 QY 464 AGCTCATCGGATATATGTCGCTACAAAAGAGAAATCAATCGGCAAAAAACAAAGACATAG 523
 Db |||||
 QY 421 TCAACATCGTGAAGGTGGAACCTCCAAGAAACGATGCTGCTGTGGCTTGGCAGCTTCGCA 480
 Db |||||
 QY 524 TCAGCATCGTGAAGGAGGACTTCAGCAAAACGATGCTGCGGTAGGCTTTGCACGTTTACA 583
 Db |||||
 QY 481 AGGACGCTTACTACAGATGATGTTTATCTTAAATGCTTCTTGATATCATAGAGATAC 540
 Db |||||
 QY 584 GGGAGCTTACACCACAGATGATGTTTATCTTCAATCACTGATATCATCAGAGATAC 643
 Db |||||
 QY 541 TGGTATGCTTATATCGTTTCTCATGGAGATCATTAACATTAATTCCTAAGATGATT 600
 Db |||||
 QY 644 GGGCGATGCTTATATCGTTTCTCATGGAGATCATTAACATTAATTCCTAAGATGATT 703
 Db |||||
 QY 601 ATCAGTAGCGAGTTGGCTGCTGCAGAGAGCCTTCTCTATCTGTCGAGAAATCTGTCAA 660

Db 704 ATCAGCTAGCGAGTGGCTGCTCAGAGAGCTTTCCCTATCTGGTCGGGAAATCTGTCAA 763
QY 661 TTCAAGAACCTATCGCGGACAAATAGCGATAACATTTCAAGAACAACTGGTACTTTC 720
Db 764 TTAAAGAACCTATCGCGGACAAATAGCGATAACATTTCAAGAACAACTGGTACTTTC 823
QY 721 TGTAAAGCAATCCAGGAATCAAAATCTAACAAGCAACAAAGCAACAACTAACAGTCA 780
Db 824 TGTAAAGCAATCCAGGAATCAAAATCTAACAAGCAACAAAGCAACAACTAACAGTCA 883
QY 781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGTAAACAGCTCTACAAATCTGG 840
Db 884 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGTAAACAGCTCTACAAATCTGG 943
QY 841 TCAACGACATGTAGATCTGATGGCTTGTCTTTGTATCCAGCACAAATCAAGTCAAG 900
Db 944 TCAACGACATGTAGATCTGATGGCTTGTCTTTGTATCCAGCACAAATCAAGTCAAG 1003
QY 901 AGCTAGAGGTGTTGCAGTGCACACGAGATCAATACCACTTCATCCCTTACTCTCAAT 960
Db 1004 CGCCAGAGGTGAGTGTCCCTCATGTTAACCATTACCACTTTATCCCTTATGAACAA 1063
QY 961 GTCTGAATCGGAAGAACGAATCGCTCGTATATTTCCCTTCGTTATCGTTCAAAACCAT 1020
Db 1064 GTCTGAATCGGAAGAACGAATCGCTCGTATATTTCCCTTCGTTATCGTTCAAAACCAT 1123
QY 1021 GGTACAGATTCAGGCGGACGAACCAACCAAGTCCACACGACCTCCGAACTAGTCCAG 1080
Db 1124 GGTACAGATTCAGGCGGACGAACCAACCAAGTCCACACGACCTCCGAACTAGTCCAG 1183
QY 1081 CCCGCACTCGACCAAACTTTAAATAGACTCAA-----ATTCTTCTTTGGT 1128
Db 1184 TCCGCACTCGACCAAACTTTCAACAGCTCCAGCAATCCAAATTCATGAGAAATGGT 1243
QY 1129 TAGTCAGCTGGTACGAAGAGTTGGGAGAGATATGTTATTTCCCTTCGTTATCGTTCAAA 1188
Db 1244 CAAAGAAAGCTGTTGCAAGAGTAGCGATGGTTATGTTCTTTGAGGAGAGATGGAGTT 1303
QY 1189 TTATGCTTTGCGAAGATTTACCATCTGAACTGTTTAAATACTTCAAGCAAGTTATC 1248
Db 1304 TTATATCCGACCAAGAACTTTTCAGCAGAAACAGCAGCAGGCAATGATGCAAACTGG 1363
QY 1249 AAAACAAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAATGTTGCTCTCGTGA 1308
Db 1364 CAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTCAGCTCCCATCTAGTGA 1423
QY 1309 CCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTG 1368
Db 1424 TCGAAGATTTTACAAATAAGGCTTATGACTTACTAGCAAGAAATCCCAAGATTTACT 1483
QY 1369 AAATAAGGCTGTAATTTCTGATTTTCAAGCCTTAGACAAATTTATAGAACGCTTGAAT 1428
Db 1484 TAATAAGGCTGCAAGATTTGATTTTGGCTTTGGATTAACCTGTTGAAACGACTCAAG 1543
QY 1429 TGAATCGACTAATAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAAT 1488
Db 1544 TGTCTCAAGTATAAAGTCAAGTTAGTGTATGATTTCTTCTGCTTCTTAGCTCCGAT 1603
QY 1489 CCATCCAGAGGCTTGGCAACCAAAATTCAAATTTGAGTATACTAGACGAGATTCG 1548
Db 1604 TCATCCAGAGGCTTGGCAACCAAAATTTGGCAAAATTTACCTACACTGATGATGAT 1663
QY 1549 TATTGCTCAATTTAGCTGATAGTATACAACGCTCAGATGGTTTACATTTTGTATGAAC 1608
Db 1664 AGTAGCAAGTTGGCAGCAGATACACACAGAGAGCGTTATATCTTTGATCCCTG 1723
QY 1609 TATAATCATGATGAAGAGATGATATGTAACCGCTCATATGGGCCATAGTCTAGT 1668
Db 1724 TATAACCAAGTATGAGGGGATGCTATGTAACCTCCACATATGACCCATAGCCACT 1783
QY 1669 TGGAAAGATAGCCTTTCTGATAGGAAAAGTTGCAAGCTCATATAAGAAAA 1728

Db 1784 TAAAAAGATAGTTTTGTGTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTTAAAGAA 1843
QY 1729 AGGTATCTCTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTC 1788
Db 1844 AGTTTGAACCCCTCTTCGACAGACCATCAGGATTCAGGAATACTGAGCAAAAGGAGC 1903
QY 1789 AGCAGCTATTTCATCAATCGTGTGAAGGGGAAAACGAATTTCCACTGCTGCTTCATTA 1848
Db 1904 AGAAGCTATCTACAACCGCGTGAAGCAGCTAAGAAAGGTGCCACTTGTATGCTTGCCT 1963
QY 1849 TATGTTGAGCATACAGTTGAGGTAAACCGTAAATTTGATTTATCTCTCATAGGATCA 1908
Db 1964 CAATCTCAATATACCTAGTAAAGTCAAAACCGGTAGTTTAAATCATACCTCTTATG 2023
QY 1909 TTACCATATATTAATTTGCTTTGGTTTGTATGATCAACATACAAAGCTCCAAATGGCTA 1968
Db 2024 TTACCATATCAATTTGAGTGGTTTGAAGAGGCTTTTATGAGSCACCTTAAGGGGTA 2083
QY 1969 TACCTTGGAAAGATTTGTTTGGCAGGATTAAGTACTAGTAGAACAACCTGACGAACTCC 2028
Db 2084 TACTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAACTCC 2143
QY 2029 ACATTTCTAATGATGGTGGGCAATCCAGTGAAGCATGTTAGGCAAGAAAGACACAG 2088
Db 2144 GCATTCAGATAATGTTTGGTAACTGCTAGCAGCATGTTCAAGAAACAAATATGCTCA 2203
QY 2089 TGAAGATCCAAATAAGAACTTTCAAAGCGATGAAGAG----- 2125
Db 2204 AGCTGATACCAATCAACGGGAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGA 2263
QY 2126 ----- 2125
Db 2264 GGAAGAAACCCCTCGAGAGAGAAAACCAAGCGAGAAAACAGAGTCTCCAAACCAAC 2323
QY 2126 -----CCAGTAGAGAAAACACTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAA 2178
Db 2324 AGAGGAACCAAGAGAGATCCAGAGGAAATCAGAGAACTCAGTCCGAGACTGAAAA 2383
QY 2179 AGTAGAGCCCAACTCAAAGAACGAGAGTTTTCGTTGCGAAAAGTAAACGAGTCTAGTCT 2238
Db 2384 GGTGAAGAAAACCTGAGAGAGCTGAAGATTTTACTTGGAAAATCCAGGATCCAAATAT 2443
QY 2239 GAAAGCAATGCAACAGAAACTCTAGCTGTTTACGAAATTAATTTGACTCTTCAAAATAT 2298
Db 2444 CAAGTCCAAATGCCAAAGAGACTCTCAGAGATTAATAAATAATTTACTATTGCGCCCA 2503
QY 2299 GGATAACAATAGTATCATGCGAAGCAGAAAATTTACTTGCCTTGTATAAGGAAGTAA 2358
Db 2504 GGACAACAATACTATTATGCGAGAGCTGAAAACACTATTGCTTTATTAAAGGAGATTA 2563

RESULT 11

ABK15103
ID ABK15103 standard; DNA; 2647 BP.

XX ABK15103;

AC ABK15103;

XX 08-MAY-2002 (first entry)

XX DNA encoding Streptococcus pneumoniae BVH-11.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;

XX streptococcal bacterial infection; gene; ds.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers

XX CDS 45..2567

XX /product= a

XX /note= "BVH-11"

XX SP64, no information on which is given in the

XX specification"

XX PN WO200198334-A2.
 XX PD 27-DEC-2001.
 XX PF 19-JUN-2001; 2001WO-CA000908.
 XX PR 20-JUN-2000; 2000US-0212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.
 XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX DR WPI; 2002-122272/16.
 XX DR P-PSDB; AAU75933.
 XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
 XX PT polypeptides, useful as vaccine components for treating or preventing
 XX PT streptococcal infections such as otitis media, meningitis, and
 XX PT bacteremia.
 XX PS Disclosure; Fig 4; 113pp; English.
 XX CC The invention describes an isolated polypeptide (I) with 70-90% identity
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
 CC Streptococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumonia protein BVH-11, used to create the antigenic
 CC peptides described in the method of the invention
 XX SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;

Query Match 57.5%; Score 1374.2; DB 6; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60
 Db |||||
 QY 104 TCGTTATGAATAGTTTGCATCAGCTCAACTGTAAAGAAATAATTCGTGTTTCCTA 163
 Db |||||
 QY 61 TATAGATGAAACAAAGCAGCGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
 Db |||||
 QY 164 TATAGATGAAACAAAGCAGCGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA 223
 Db |||||
 QY 121 CGGTGAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
 Db |||||
 QY 224 CGGTGAGGAATCAAGCGCGCAAAATCGTCATCAAGATAACAGGATTCAGGATTCATGTGAC 283
 Db |||||
 QY 181 TTCAATGCGGACCACTATCATTTATFACAAATGTAAGGTTCCCTATGACGCTATCATCAG 240
 Db |||||
 QY 284 CTCTCATGAGACCATTAATCATTAATGTAAGGTTCCCTATGATGCCATCATCAG 343
 Db |||||
 QY 241 TGAAGAATTAATCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATGTTAATGA 300
 Db |||||
 QY 344 TGAAGAGTCTCTCATGAAAGATCCGAATTAATCATGTTGAAGATTCAGACATTCATGA 403
 Db |||||
 QY 301 GGTCAAGGTTGATATGTTTCAAGGTAGATGAAATACTATGTTTACCTTAAGATGC 360
 Db |||||
 QY 404 AATCAAGGTTGTTATGTTCAITTAAGGTTAAACGGTAAATACTATGTTTACCTTAAGGATGC 463
 Db |||||

QY 361 TGCCACCGCGGATAACGTCCGTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420
 Db |||||
 QY 464 AGCTCATGCGGATAATGTCCGTACAAAGAGGAAATCAATTCGCGCAAAACAAAGACATAG 523
 Db |||||
 QY 421 TCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGTTGCCCTTGCACGTTCCGA 480
 Db |||||
 QY 524 TCAGCATCGTGAAGGAGGAGACTTCAGAAACGATGGTGGGTAGCCCTTTCGACGTTTACA 583
 Db |||||
 QY 481 AGGACGCTATPACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGATAC 540
 Db |||||
 QY 584 GGGACGCTACACACAGATGATGTTTATATCTTCAATGATCTGATATCATCGAAGATAC 643
 Db |||||
 QY 541 TGGTGATGCTTATATGTTCTCTCATGAGATCATTAACCATTAATTCCTTAAGATGATTT 600
 Db |||||
 QY 644 GGGCGATGCTTATATGTTCTCTCATGAGATCATTAACCATTAATTCCTTAAGATGATTT 703
 Db |||||
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAGCTTCCCTATCTGCTGAGGAAATCTGTCAA 660
 Db |||||
 QY 704 ATCAGCTAGCGAGTTGGCTGCTGAGAGCTTCCCTATCTGCTGCGGAAATCTGTCAA 763
 Db |||||
 QY 661 TTCAGAACCTATFCGCGCAAAATAGCGATAACCTTCAAGAACAACTTGGGTACCTTC 720
 Db |||||
 QY 764 TTTAAGAACCTATFCGCGCAAAATAGCGATAACCTTCAAGAACAACTTGGGTACCTTC 823
 Db |||||
 QY 721 TGTAAAGCAATCCAGGAGCTACAAATTAACAAAGCAACCAAGCAACCACTTAACAGTCA 780
 Db |||||
 QY 824 TGTAAAGCAATCCAGGAGCTACAAATTAACAAAGCAACCAAGCAACCACTTAACAGTCA 883
 Db |||||
 QY 781 AGCAAGTCAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 Db |||||
 QY 884 AGCAAGTCAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 943
 Db |||||
 QY 841 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 Db |||||
 QY 944 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1003
 Db |||||
 QY 901 AGCTAGAGGTTGTCAGTGCACACGAGATCATTAACCATTAATTCCTTACTCTCTCAAT 960
 Db |||||
 QY 1004 CGCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
 Db |||||
 QY 961 GTCTGAATTTGGAAGAACGAATCGCTGATTAATTCCTTCTGTTTATGTTTCAACCATG 1020
 Db |||||
 QY 1064 GTCTGAATTTGGAAGAACGAATTCGCTGATTAATTCCTTCTGTTTATGTTTCAACCATG 1123
 Db |||||
 QY 1021 GGTACAGATTCAGGCCAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCCAG 1080
 Db |||||
 QY 1124 GGTACAGATTCAGGCCAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCCAG 1183
 Db |||||
 QY 1081 CCGCAACCTGCAACCAATCTTAAATAGACTCAA-----ATTCTCTTTGGT 1128
 Db |||||
 QY 1184 TCGCAACCTGCAACCAATCTTAAATAGACTCAA-----ATTCTCTTTGGT 1243
 Db |||||
 QY 1129 TAGTCAGCTGGTACGAAAGTTGGGGAAGGATATGTTTCGGAAGAAAGGGCATCTCTCG 1188
 Db |||||
 QY 1244 CAAAGAGCTGTTCCGAAAGTAGGCGATGTTTATGTTTTCGTTGAGGAGATGAGGATTTCTCG 1303
 Db |||||
 QY 1189 TTATGCTTTTGGAAAGATTTTACCATCTGAAATCTGTTTAAATCTTTGAAAGCAAGTTATC 1248
 Db |||||
 QY 1304 TTATATCCGACCAAGAAATCTTTCAGCAGAAACAGCAGCAGCAATGATAGCAAACTGGC 1363
 Db |||||
 QY 1249 AAAACAGAGAGTGTTCACACATTTTAACTGCTTAAAGAAAGAAATGTTGCTCCCTGTA 1308
 Db |||||
 QY 1364 CAAGCAGGAAAGTTTATCTCAAGCTAGGAGCTAAGAAACCTGACCTCCCTCTAGTGA 1423
 Db |||||
 QY 1309 CCAAGATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTATAAGGCTCTTTGN 1368
 Db |||||
 QY 1424 TCGAATTTTACATAAGGCTTATGATTTACTAGCAAGATTTCCAAAGATTTACTTGA 1483
 Db |||||
 QY 1369 AAATAAGGTCGTAAATCTGATTTTCCAAAGCTTTAGACAAATTTATTAGAACGCTTGAATGA 1428
 Db |||||
 QY 1484 TAATAAGGTCGACAAAGTTGATTTTGGGCTTTGGATTAACCTGTTTGGAAACGACTCAAGA 1543
 Db |||||

QY 298 TGAGGTCAAGGTGGATATGTTTCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 357
DB 360 TGAATCAAGGTGGTATGTTTCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 419
QY 358 TGCTGCCACCGGGATAACGTCGTAACAAAGAGGAAATCAATCGACAAACAAAGAGCA 417
DB 420 TGCAGCTCATCGGATAATATTCGGACAAAGAGAGATTAAACGTCAGAAGCAGGAACA 479
QY 418 TAGTCACATCGTGAAGGTGGAACTCAAGAAACGATGGTGTGCTGTCCTTGGCCTTGGCAGCTTC 477
DB 480 CAGTCATAATCAACGGGGGTGGTCT-----AACGATCAAGCAGTAGTTGCGACCCAGAGC 533
QY 478 GCAAGCAGCTTACTACAGATGATGGTTATATCTTTTAATGCTTCGATATATCATAGGA 537
DB 534 CCAAGCAGCTATACACCGATGATGGTTATATCTTCAATGCACTGATATCATTCAGGA 593
QY 538 TACTGGTATGCTTATATATCGTTCCTCATGGAGATCAATACCATTACATTCCTAAGGAATGA 597
DB 594 CACGGGTGATGCTTATATCGTTCCTCACGGCAGCAATACCATTACATTCCTAAGGAATGA 653
QY 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCCTATCTGGTCGAGGAATCTGTC 657
DB 654 GTTATCAGCTAGCGAGTTAGCTGCTGCAGAACCTTATGGAAATGG----- 698
QY 658 AAATTCAGAACTATCGCGCAGCAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACC 717
DB 699 -GAAGCAGGATCTCGTCTCTTCAAGTTCCTAGTTTATATGCAATTCAGCTCAACCAA 757
QY 718 TTCTGTGAAGCAATCCAGGAACTACAAATATAACACAGCAACACAGCAACACTAACAG 777
DB 758 GATTGTGAGAGAACCAATCT-----GACTGTCACTCCAACTTA 797
QY 778 TCAAGCAGCTCAAGTAATGACATTGATAGTCTCTTGAACAGCTCTACAAACTGCTTT 837
DB 798 TCATCAAAATCAAGGGGAAACATTTCAAGCCCTTTACGTGAATGTATGCTAAACCCCTT 857
QY 838 GAGTCAACGACATGTAGAACTCTGATGCCCTTGCTTTGATTCAGACACAAATCACAGTCG 897
DB 858 ATCAGAACGCCATGTGGAACTCTGATGGCCCTTATTTTCGACCCAGCGCAAAATCACAACTCG 917
QY 898 AACAGCTAGAGGTGTGCAAGTGCCACACGAGATCAATTACCACCTTCATCCCTTACTCTCA 957
DB 918 AACCGCCAGAGGTAGTGTCCCTCATGTAAACCAATACCACCTTTATCCCTTATGAACA 977
QY 958 AATGCTGAATTTGGAAGAACGATCGCTGTATATTCCTCTGTTATCGTTTCAAAACCA 1017
DB 978 AATGCTGAATTTGGAAGAACGATTCGTTATATTCCTCTGTTATCGTTTCAAAACCA 1037
QY 1018 TTGGGTACCAGATTCAGGGCCAGAACCAACAACTCCACACCGACTCCGGAACTAGTCC 1077
DB 1038 TTGGGTACCAGATTCAGAACCCAGAACCAACAACTCCACCACTCCGAACTAGTCC 1097
QY 1078 AGGCGCGAACTGCAACCAATCTTAAATAAGACTCAA-----ATTCTTCTTT 1125
DB 1098 AAGTCGGCAACCTGCAACCAATCTTCAACAGCTCCAGCAATCCCAATTTGATGAGAAAT 1157
QY 1126 GGTAGTCACTGCTACGAAAGTTGGGGAAGGATATGTTTCAAGAAAGGCACTCTC 1185
DB 1158 GGTCAAGAGAGCTGTTTCGAAAGTAGCGGATGTTATGCTTTGAGGAGAATGGAGTTTC 1217
QY 1186 TCGTTATGCTTTTGGCAAGATTTACCATCTGAACTGTTTAAATACTTTGAAAGCAAGTT 1245
DB 1218 TCGTTATATCCAGCCAGCAAGATCTTTCAGCAGAACACAGCAGGCAATTTGATAGCAACT 1277
QY 1246 ATCAAAACAGAGAGTGTTCACACTTTTAACTGCTTAAAGAAAGAAATGTTGCTCCTCG 1305
DB 1278 GGCCAAAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAAACTGACCTCCCACTAG 1337
QY 1306 TGACCAAGAAATTTTATGATAAAGCATATAATCTGTTTAACTGAGGCTCATTAAGCCCTGTT 1365
DB 1338 TGATCGAGAAATTTTAAATAAGSCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
QY 1366 TGNAAATAAGGTCGTAATTTCTGATTTTCAAGCCTTTAGACAAATATTATAGACGCTTGAA 1425

DB 1398 TGATATAAAGTTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAACGACTCAA 1457
QY 1426 TGATGAATCGACTAATAAAGAAAAATTTGGTAGATGATTTATTGGCAATCTCTAGACCAAT 1485
DB 1458 GGATGTCCTCAAGTGAATAAGTCAAGTTAGTGGATGATATTCCTTGGCTTCTTAGCTCGAT 1517
QY 1486 TACCCATCCAGAGCAGCTTTGGCAACCAATTTCTCAATTTAGTATATCTAGAGAGCAAGT 1545
DB 1518 TCGTCATCCAGAACGTTTAGGAAAAACCAATCGCAAAATTTACCTACACTGATGATGAGAT 1577
QY 1546 TCGTATTTGCTCAATTAGCTGATAAGTATACAAGTCAACGTCAGATGGTTACATTTTGGATGAACA 1605
DB 1578 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAACGGTTATATCTTTGATCCTCG 1637
QY 1606 TGATATAATCAGTGAAGAGAGATGCATATGTAAGCGCTCATATGGGCCATAGTCACTG 1665
DB 1638 TGATATAACCACTGATGAGGGGATGCTTATGTAACCTCCACATATGACCCATAGCCACTG 1697
QY 1666 GATTGGMAAGATAGCTTTCTGTAAAGAAAAAGTTGCAGCTCAAGCCTATCTACTAAAGA 1725
DB 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCCAGGCTTTATGCTAAAGA 1757
QY 1726 AAAAGTATCTTACCTCCATCTCCAGACGACATGTTAAAGCAAAATCCAACTGGAGATAG 1785
DB 1758 GAAAGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGAAATACTAGAGCAAAAG 1817
QY 1786 TCGACGAGCTATTTACAATCGTGTGAAAGGGAAAAAGAAATTCACCTCGTTTCGACTTCC 1845
DB 1818 AGCAGAAAGCTATCTCAACCGCTGAAAGCAGTAAAGAGGTGCCACTTTGATCGTATGCC 1877
QY 1846 ATATATGGTTGAGCATACAGTTGAGTTTAAACCGTAAATTTGATTTATTTCTCATAGGA 1905
DB 1878 TTACAATCTTCAATATACTGTAGAAGTCAAAAACGAGTAGTTTAACTCATCTCATATGA 1937
QY 1906 TCATTACCAATAATATAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGG 1965
DB 1938 CCAATTACCAATACATCAAAATTTGAGTGGTTTGAAGAGCCTTTATGAGGACCTTAAGG 1997
QY 1966 CTATACCTTTGGAAGATTTGTTTGCAGCAGTTAAGTACTACGTAGAACACCTTCAGCAAGC 2025
DB 1998 GTATATCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAAG 2057
QY 2026 TCCACATCTTATGATGATGGGCAATGCCAGTAGCAGCTGTTAGGCAAGAGAACCA 2085
DB 2058 TCCGCAATTCAGATAATGTTTGGTAAACGCTAGCAGCACTATGTTCAAAAGAAACAAAAATGG 2117
QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGA----- 2121
DB 2118 TCAAGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2177
QY 2122 ----- 2121
DB 2178 TGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAAACCCAGAGTCTCCAAAAACC 2237
QY 2122 ---AGAGCCAGTAGAGGAAACCTCTGAGCCAGAGTCCCTCAAGTAGACACTGAAAA 2178
DB 2238 AACAGAGAACCCAGAGAAATCCAGAGGAATCCAGAGAACCTCAGGTCGAGACTGAAAA 2297
QY 2179 AGTAGAAGCCCAACTCAAAGAAAGAGAGTTTGTCTCGAAAGTAAACGAGTTCTAGTCT 2238
DB 2298 GGTTCAGAAAAACTGAGAGAGCTGAGATTTACTTCGAAAAATCCAGGATCCAAATAT 2357
QY 2239 GAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAATTTGACTTTCAAATAT 2298
DB 2358 CAACTCCAAATGCAAGAGACTCTCACAGGATTTAAAAAATAATTTTACTATTTGGCACCCA 2417
QY 2299 GATTAACATAGTATCATGGCAGAGACGAGAAAAATTTCTTGGTTGTTTAAAGAGAGTAA 2358
DB 2418 GGACAAACAATATATTATGGCAGAGAGCTGAAAAAATATTGGCTTTATTAAAGAGAGATAA 2477

AAA05417	QY	238	CAGTGAAGAATTA	297
ID AAA05417 standard; DNA; 2481 BP.				
XX	Db	300	CAGTGAAGAGCT	359
AC AAA05417;				
XX				
DT 24-MAY-2000 (first entry)	QY	298	TGAGGTCAAGGGT	357
XX	Db	360	TGAAATCAAGGGT	419
DE Streptococcus pneumoniae nucleotide sequence ID311.				
XX	QY	358	TGCTGCCACGCG	417
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;				
KW antibacterial; antinflammatory; meningitis; infection; diagnosis;	Db	420	TGAGGTCAAGGGT	479
XX pneumococcal disease; ds.				
XX	QY	418	TAGTCAACATCG	477
OS Streptococcus pneumoniae.	Db	480	CAGTCATAATCA	533
XX				
PN W0200006737-A2.	QY	478	GCAGGACGCTAT	537
XX	Db	534	CCAAGGACGCTA	593
PF 10-FEB-2000.				
XX	QY	538	TACTGGTGATGT	597
PF 27-JUL-1999; 99WO-GB002451.	Db	594	CACGGGTGATGT	653
XX				
PR 27-JUL-1998; 98GB-00016337.	QY	598	GTTATCAGCTAG	657
PR 19-MAR-1999; 99US-0125164P.	Db	654	GTTATCAGCTAG	698
XX				
PA (MICR-) MICROBIAL TECHNIQS LTD.	QY	658	AAATTCAGAACCT	717
XX	Db	699	-GAAGCAGGGAT	757
PI Gilbert CFG, Hansbro PM;				
XX	QY	718	TTCTGTAAGCAAT	777
DR WPI; 2000-195300/17.	Db	758	GATTGTCAGAGAC	797
DR P-PSDB; AAY81662.				
XX	QY	778	TCAGCAAGTCAA	837
XX	Db	798	TCATCAAAATCA	857
CC New Streptococcal protein, useful as a vaccine, for diagnosis of				
CC pneumococcal diseases and for screening agents capable of antagonizing or	QY	838	GAGTCAACGACAT	897
PT inhibiting expression of the protein.	Db	858	ATCAGAACGCGAT	917
XX				
PS Claim 2; Page 99; 108pp; English.	QY	898	AAACAGCTAGAGT	957
XX	Db	918	AAACGCGCAGAGT	977
CC AAY81501 to AAY81679 represent specifically claimed protein sequences				
CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent	QY	958	AATGTCGTAAT	1017
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.	Db	978	AATGTCGTAAT	1037
CC The sequences have antibacterial and antiinflammatory properties. The				
CC protein sequences, and fragments of them, are useful as immunogens and/or	QY	1018	TTGGGTACCAAGT	1077
CC antigens. The nucleotide sequences can be used in vaccines and in	Db	1038	TTGGGTACCAAGT	1097
CC diagnostic assays. The proteins and nucleotides can be useful for the				
CC detection and diagnosis of S. pneumoniae. The protein sequences are also	QY	1078	AGGCCCGCAACT	1125
CC useful for screening an agent capable of antagonising, inhibiting or	Db	1098	AAGTCCGCAACT	1157
CC interfering with the function or expression of the proteins in which the				
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection	QY	1126	GTTTAGTACGTGT	1185
CC and meningitis. AA05591 to AA05614 represent primers used in the	Db	1158	GGTCAAGAGAGCT	1217
CC exemplification of the present invention				
XX	QY	1186	TCGTTATGTCCT	1245
SQ Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 U; 0 Other;	Db	1218	TCGTTATGTCCT	1277
Query Match 42.0%; Score 1003.8; DB 3; Length 2481;	QY	1246	ATCAAAAACAGAG	1305
Best Local Similarity 66.0%; Pred. No. 4e-243;	Db	1278	GGCCCAACGAGAA	1337
Matches 1623; Conservative 0; Mismatches 693; Indels 144; Gaps 6;				
	QY	1	TTCTACGAGTTGG	57
	Db	60	TTCTATGAGCTTG	119
	QY	58	CTATATAGATGGA	117
	Db	120	TTATATAGATGTT	179
	QY	118	CAACGCTGAAGGA	177
	Db	180	TAAGAGGAGGGAT	239
	QY	178	CACCTTCATCGGC	237
	Db	240	GACCTTCATCGGC	299

Db 1338 TGATCGAGATTTTACAAATAGGCTTATGCTTACTAGCAAGATTCACCAAGATTACT 1397
 QY 1366 TGNAAATAGGGTGTGAATCTCTGATTTTCCAGCTTAGCAAAATATATAGAACGCTTGAA 1425
 Db 1398 TGATAATAAGGTCGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAACGACTCAA 1457
 QY 1426 TGATGAATCGATTAATAAGAAAAATTTGATAGATGATTTATTTGGGATTCCTAGCACCAAT 1485
 Db 1458 GGATGTCCTCAAGTGATAAAGTCAAGTTAGTGATGATATCTTTGCCCTCTTAGCTCCGAT 1517
 QY 1486 TACCATCCAGACGCTTTGGCAACCAAAATTTCTCAAATTTGAGTATPACTGGAAGCAAGT 1545
 Db 1518 TCGTCAATCCAGAACGTTTAGGAAAAACCAATGCGCAAAATTTACTACACTGATGAGAT 1577
 QY 1546 TCGTATGCTCAATAGCTGATAAGTATACAGCTGAGATGTTTATCTTTTATGTAACA 1605
 Db 1578 TCAAGTAGCCTAAGTTGGCAGGCAAGTACACAGAAAGAGGTTATATCTTTGATCCTCG 1637
 QY 1606 TGATATAATCAGTGATGAAGGAGATGCATATGTAAGCGCTCATATGGGCCATAGTCACTG 1665
 Db 1638 TGATATACCAAGTATGAGGGGATGCTTATGTAATCTCACATATGACCATAGCCACTG 1697
 QY 1666 GATTGGAAGAAGATAGCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTATATACTAAAGA 1725
 Db 1698 GATTAAAAAAGATAGTTTCTGTAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAAGA 1757
 QY 1726 AAAAGGTATCTACCTCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAG 1785
 Db 1758 GAAAGTTTGGACCTCTCTTCACAGACCATCAGGATTCAGGAAAAATCTAGGAGCAAAAG 1817
 QY 1786 TCAGCAGCTATTTCAATCGTGTGAAGGGAAAAACGAAATTCACCTCGTTGCACTCC 1845
 Db 1818 AGCAAGAGCTATCTACACCGGCTGAAAGCAGCTAAGAGGTGCCACTTGATCGTATGCC 1877
 QY 1846 ATATATGTTGAGCTACATAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAAGGA 1905
 Db 1878 TTACAATCTTCAATATCTGTAAGAGTCAAAACGGTAGTTTAAATCATACCTCATATGA 1937
 QY 1906 TCATTACCATATATTAATTTGCTGTTGATGATCACATACAAAGCTCCAAATGG 1965
 Db 1938 CATTACCATACATCAATCAATTTGAGTGGTTTTCAGGAAGGCTTTTATGAGGCACTAAAGG 1997
 QY 1966 CTATACCTTGGAGATTTGTTTGGCAGGATTAAGTACTAGTAGAACACCTGAGCAAGC 2025
 Db 1998 GTATCTCTGAGATCTTTTGGCAGCTGCAAGTACTATGTCGAACATCTCAAGCAAGC 2057
 QY 2026 TCCATCTTAATGATGAGTGGGCAATCCAGTGAGCATGTTGTTAGGCAAGAAAGACCA 2085
 Db 2058 TCCGCATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAGAAACCAAAATGG 2117
 QY 2086 CAGTGAAGTCAAAATAGAACTTCAAGCGATGA----- 2121
 Db 2118 TCAAGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAACCC 2177
 QY 2122 ----- 2121
 Db 2178 TGAGGAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAACCC 2237
 QY 2122 ---AGAGCCAGTGAAGAAACACCTGTCGAGCGAGAGTCCCTCAAGTAGAGACTGA AAA 2178
 Db 2238 AACAGAGAAACCGAAGAAATCACAGAGGAATCAGAGAAACCTCAGGTCGAGACTGAAA 2297
 QY 2179 ACTAGAGCCCACTCAAGAGAGCAGAAAGTTTGGCTTGGCAAGTAAAGGATCTAGTCT 2238
 Db 2298 GGTGGAAGAAACCTGAGAGAGGCTGAAGATTTACTTTGGAAAAATCCAGAGTCCAAATTAT 2357
 QY 2239 GAAAGCCTATGCAACAGAACTCTAGCTGTTTACGAAATAATTTGACTCTTCAATAT 2298
 Db 2358 CAAGTCCANTGCAAGAGACTCTACAGATTTAAAAATATTTTACTATTTGGCACCCA 2417
 QY 2299 GGATAACATATGATCATCGCAGAGACAGAAAAATTTACTTGGCTTGTGTTAAAAAGGAGTAA 2358

Db 2418 GGACAACRATACTATTATGGCAGAGACTGAAAAACTATTGGCTTTATTAAAGAGAGTAA 2477
 RESULT 14
 ABX06885
 ID ABX06885 standard; DNA; 2457 BP.
 XX AC ABX06885;
 XX DT 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX S. pneumoniae type 4 strain coding region #1173.
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX Streptococcus pneumoniae; type 4 strain.
 OS WO200277021-A2.
 PN 03-OCT-2002.
 XX 27-MAR-2002; 2002WO-IB002163.
 XX 27-MAR-2001; 2001GB-00007659.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR P-PSDB; ABU01597.
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX Claim 6; SEQ ID NO 2345; 56pp; English.
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2489
 CC identified coding region from the genomic sequence. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)

XX	Sequence	2457 BP; 836 A; 497 C; 531 G; 593 T; 0 U; 0 Other;	
SQ	Query Match	41.6%; Score 993; DB 7; Length 2457;	
	Best Local Similarity	65.8%; Pred. No. 2.1e-240;	
	Matches 1611; Conservative	0; Mismatches 696; Indels 141; Gaps 6;	
Qy	1	TTCTTACAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC	57
Db	60	TTCTTATAGCTTGGAGCTTACCAAGCTGGTCAGATAAGAAAGTCTAATCAGTTGC	119
Qy	58	CTATATAGATGGAAAAACAAGCGACGCAAAAACCGAGAAATTTGACTCCTCGATGAGTTAG	117
Db	120	TTATATAGATGGTGCATCAGCTGTCTAAAGGCGAGAAAACCTTGACACAGATGAATCAG	179
Qy	118	CAAGCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCTATGT	177
Db	180	TAAGAGGAGGGGATCAACGCGAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT	239
Qy	178	CACCTTACATGGGACCACTATCATTTATTAATGAGTAAAGTTCCTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGACCAATTAATCAATTAATGAGCAAGTTCCTTATGATGCCATCAT	299
Qy	238	CAGTGAAGAAATTACTCATGAAGAATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGAATCCGAATTAATCAGTTGAAGGATTCAGACATTTGCAA	359
Qy	298	TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAATTAATTAATTAATTAAGGA	357
Db	360	TGAATCAAGGGTGGTATGTCATTAAGGTAAACGGTAAATTAATTAATTAATTAAGGA	419
Qy	358	TGCTGCCACGCGATTAAGCTGCTGACAAAGAGAAATCAATCGACAAAACCAAGAGCA	417
Db	420	TGCAAGTCAATGCGGATTAATTCGACAAAGAGAGATTAAGCTCAGAGCAGGAAAG	479
Qy	418	TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGGTGTCTGCTTGGCACGTTTC	477
Db	480	CAGTCATAATCAT-----AACTCAAGAGCAGATAATGCTGTGCTGCAGCCAGAGC	530
Qy	478	GCAGGAGCGCTATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGA	537
Db	531	CCAGGAGCGTTATACAAACGGATGATGGGTATATCTCAATGCTATCATATTTAGAGGA	590
Qy	538	TACTGTGTGATCTTATATCGTTCCTCATGGAGATCAATACCATTAATTCCTAAGAAATGA	597
Db	591	CACGGGTGATGCTTATATCGTTCCTCACGGGACCAATTAATTAATTAATTAATTAAGGA	650
Qy	598	GTTATCAGCTAGCGATGGTGTGTCAGAGAGCTTCTTATCTGCTGAGGAAATCTGTC	657
Db	651	GTTATCAGCTAGCGATGGTGTGTCAGAGAGCTTCTTATCTGCTGAGGAAATCTGTC	695
Qy	658	AAATTCAGAGACCTATCGCGACAAATAGCGATAACACTCAAGAAACAACTGGGTACC	717
Db	696	-GAAGAGGAGATCTCGTCTTCTTCAAGTTCATGTTAATGCAAAATCCAGCTCAACCAA	754
Qy	718	TTCTGTAAAGCAATCCAGGAACTACAAATTAATTAACAAAGCAACCAAGCAACATAACAG	777
Db	755	GATTTGTCAGAGAACCAATCT-----GACTGTCACTCCAACTTA	794
Qy	778	TCAAGCAAGTCAAAAGTAAATGATGATGATGCTCTTGAAGACAGCTCTACAAATCGCTTT	837
Db	795	TCATCAAAATCAAGGGGAAACATTTCAAGGCTTTTACGTGAATTTGATGTAACCCCTT	854
Qy	838	GAGTCAACGACATGTAGAACTGATGCTGCTTGTGATGTCAGACCAAAATCAAGTCG	897
Db	855	ATCAGAACCGCATGTGGAATCTGATGCTTATTTTTCAGCCAGCGCAATCAAGTCG	914
Qy	898	AACAGCTAGAGGTGTGAGTGCCACAGGAGATCATTAACATTCATTCCTTACTCTCA	957
Db	915	AACGCCAGAGGTGTAGCTGTCCCTCATGTTAAACCAATTAACCACTTATCCCTTATGAACA	974
Qy	958	AATGCTGAATTGAAGAACCAATCGCTCGTATTAATTCCTTCTGTTATCGTTCAAAACCA	1017

Db	975	AATGCTCTGAATTTGAAAAACGAATTTGCTCGTATTAATTTCCCTTCTGTTATCGTTCAAAACCA	1034
Qy	1018	TTGGGTACAGATTTCAAGGCCAGAACAAACCAAGTCCACAAACCGACTCCGAACTACTGCC	1077
Db	1035	TTGGGTACAGATTTCAAGAACCAAGAACCAAGTCCACAAACCGACTCCGAACTACTGCC	1094
Qy	1078	AGGCCGCAACCTCGACCAAAATCTTAAATATAGACTCAAAATCTTCTTTTGGTGTAGTACGT	1137
Db	1095	AAGTCCGCAACAGCTCCAAGCAATCCAATTTGATGAA-----ATTGCTCAAGAGAC	1148
Qy	1138	GGTACGAAAAAGTTGGGGAAGGATATGATTTTCGAAGAAAAAGGATCTCTCTGTTATGCTT	1197
Db	1149	TGTTCCAAAAAGTAGGCGATGCTTATGCTTTTGGAGAGATGGAGTTTCTCGTTATATCCC	1208
Qy	1198	TGCGAAGATTTACCATCTGAAACTGTTAAATCTTTGAAGCAAGTTATCAAAACAAGA	1257
Db	1209	AGCCAAGGATCTTTACGAGAAAACAGCAGCGCAATTTGATAGCAAACTGGCCAGCAGGA	1268
Qy	1258	GAGTGTTCACACACTTTTAACTGCTAAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAATT	1317
Db	1269	AAGTTTATCTCATAGCTAGGAACTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATT	1328
Qy	1318	TTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATATAGGG	1377
Db	1329	TTCAATATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGTATATAAAGG	1388
Qy	1378	TCGTAAATCTGATTTCCAAAGCTTAGACAAATTAATTAAGAACGCTTGAATGATGAATCGAC	1437
Db	1389	TCGACAAAGTGTATTTGAGGCTTTGGATAACCTGTTGGAAAGCAGCTCAAGATGTTCTCAAG	1448
Qy	1438	TAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACCAATTAACCCATCCAGA	1497
Db	1449	TGATAAAGTCAAGTTAGTGAAGATATTTCTTGCCTTCTTAGCTCCGATTCGTCAATCCAGA	1508
Qy	1498	GCAGCTTGGCAAAACCAAAATTTCAAATTTAGTATGATCTAGAACAGAAAGTTTGTATGCTCA	1557
Db	1509	ACGTTTAGSAAAAACCAATTCGCAAAATTTACCTACACTGATGATGAGATTTCAAGTAGGCAA	1568
Qy	1558	ATTAGCTGTAAGTATACAAACGTCAGATGGTGTACATTTTGTGAACATGATATAATCAG	1617
Db	1569	GTGGCAGGCAAGTACACAAACAGAGAGCGTTATATCTTTGATCTCTGATATAACAG	1628
Qy	1618	TGATGAAGAGATGCTATATGTAACGCTCATATGCGCCATAGTCACTGGATTTGGAAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAATCCACATATGACCCATAGCCACTGGATTTAAAAAGA	1688
Qy	1678	TAGCCTTTCTGATGAAGAAAAAGTTGACGCTCAAGCCTATATACTAAAGAAAAAGGATTCCT	1737
Db	1689	TAGTTGCTGTAAGCTGAGAGAGCGGCGCCAGGCTTATGCTTAAAGAGAAAGGTTTGAC	1748
Qy	1738	ACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTAT	1797
Db	1749	CCCTCTTCAGACAGCATCAGGATTCAGGAAATCTAGGCAAAAGAGGAGCAGAGCTAT	1808
Qy	1798	TTACAATCTGTGTGAAGGGGAAAAACGAATTTCACTCGTTCGACTTCATATATGTTGA	1857
Db	1809	CTACAACCGCGTGAAGCAGCTAAGAAAGTGCCACTTGATGCTATGCTTACATCTTCA	1868
Qy	1858	GCATACAGTTGAGGTTAAACCGTAAATTTGATTTTCTTCAATGAAGATCATTAACCATAA	1917
Db	1869	ATATACGTGAAGAGTCAAAACCGTAGTTTAACTACCTCAITATGACCAATTAACCATAA	1928
Qy	1918	TATTAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGA	1977
Db	1929	CATCAATTTGAGTGGTTTGACGAAGGCTTTTATGAGGCACTTAAGGGTATACTCTTGA	1988
Qy	1978	AGATTGTTTGGAGCGATTAAGTACTACGTAGAAACACCTTGAAGAGCTCCACATTTCTAA	2037
Db	1989	GGATCTTTTGGCGACTGCTCAAGTACTATGTGCAACATCCAAACGAACTCCGCAATTCAGA	2048
Qy	2038	TGATGATGGGGCAATCCCAAGTGAAGTGTGTTAGGCAAGAAAGACACACAGTGAAGATCC	2097
Db	2049	TAATGGTTTGGTAACTAGCGACCAATGTTCAAGAAACAAAAATGGTCAAGCTGATAC	2108

QY 2098 AATTAAGAACTTCAAGCGGATGA----- 2121
 Db 2109 CAATCAACCGGAAACCAAGCGGAGGAGAAACCTCAGACAGAAACCTCAGGAGAAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCTCGAGNAGAAACCGCAAGCGGAGAAACAGAGTCTCCAAACCAACAGAGAAC 2228
 QY 2131 AGAGAAACACTGTGTGAGCCGAGAGTCCCTCAAGTAGAGACTGAAAGAGTGAAGCCCA 2190
 Db 2229 AGAAGAATCACCAGAGGAATCAGAGAACTCAGGTGCGAGCTGAAAGAGTTGAAGAAA 2288
 QY 2191 ACTCAAGAGCAGAGTCTTCTGCGGAGAGTAAAGGATCTAGTCTCAAGCCCAATGC 2250
 Db 2289 ACTGAGAGGCTGAAGATTTACTTTGGAAGAAATCCAGGATCCAATTTATCAAGTCCAATGC 2348
 QY 2251 AACAGAACTCTAGTGTGTTTACGAAATATTTGACTCTTCAAAATTTATGATAACAATAG 2310
 Db 2349 CAAGAGACTCTCAGAGGATTAAGAAATATTTACTATTGTCACCCAGAGCAACATAC 2408
 QY 2311 TATCATGGCAGAGCAGAAATATTTCTGCGTGTGTTTAAAGAGGATTA 2358
 Db 2409 TATTATGGCAGAGCTGAAAGAACTATTGGCTTTATTAAAGAGAGTAA 2456

RESULT 15

AAA47605
 ID AAA47605 standard; DNA; 2531 BP.
 AC AAA47605;
 XX
 DT 20-OCT-2000 (first entry)
 DE Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 OS
 XX Streptococcus pneumoniae.
 XX
 PH Location/Qualifiers
 FT CDS
 FT 1..2531
 FT /*tag= a
 FT /product= "Sp36B polypeptide"
 XX
 PN WO200037105-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US030390.
 XX
 PR 21-DEC-1998; 98US-0113048P.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Johnson LS, Koenig S, Adamou JE;
 XX
 DR WPI; 2000-452129/39.
 DR P-PSDB; AAB01469.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX
 PS Disclosure; Page 69-70; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of S. pneumoniae in order to utilize such polypeptides in vaccines to

CC protect against a wide variety of S. pneumoniae. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHx) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections

SQ Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 0 U; 3 Other;

Query Match 41.5%; Score 991; DB 3; Length 2531;
 Best Local Similarity 65.7%; Pred. No. 6.9e-240;
 Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGCACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTGTTC 57
 Db 60 TTCTTATGAGCTTGGACGTTTACCAAGCTGTCAGGATAGAAGAGTCTATCGAGTTGC 119
 QY 58 CTATATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGTTAG 117
 Db 120 TTATATAGATGCTGATCAGGCTGGTCAAAAGGCAGAAAACCTTGACACAGATGAAGTCAG 179
 QY 118 CAAGCGTGAAGCAATCAATGCTGAGCAATCTCATCAAGATAACAGACCAAGGCTATGT 177
 Db 180 TAAAGAGGGAGGGGATCAACGCCGAAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT 239
 QY 178 CACTTTCATATGGCGACCACTATCATTTATTAATGTAAGTTCCTTATGAGCTATCAT 237
 Db 240 GACCTCTCATGGAGACCATTAATCTATTAATGCGAAGTTCTTATGATGCCATCAT 299
 QY 238 CAGTGAAGATTTACTATGAAGATCCAAACTATAAGCTTAAAGATAGAGATATTGTTAA 297
 Db 300 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTGTCAA 359
 QY 298 TGAGCTCAGGCTGATATGTTATCAAGTAGATGGAATATCTATTTACTTTAAGGA 357
 Db 360 TGAATCAAGGGTGGTTATGTTCAITTAAGGTAAACGGTAAATATCTATGTTTACCTTTAAGGA 419
 QY 358 TGCTGCCCAACGGGATAAGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
 Db 420 TGCRCCTCATGGGATATATTTGCGACAAAGAGAGATTAACGTGAGACGAGAACG 479
 QY 418 TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGCTGTGCTTGGCGACGTTTC 477
 Db 480 CAGTCAATATCAT-----AACTCAAGACAGATAATGCTGTTGCTGCAGCCAGAGC 530
 QY 478 GCAAGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGA 537
 Db 531 CCAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGATCTGATATCAITTGAGA 590
 QY 538 TACTGCTGATGCTTATATCTGTTCTTCATGGAGATCAATACCATTTACATTCCTAAGAATGA 597
 Db 591 CACGGGTGATGCTTATATCTGTTCTTCACGGCGACCATTAACCATTAATCTCTAAGATGA 650
 QY 598 GTTATCAGCTAGCGAGTTGGCTGCTGAGAGAGCTTCTCTATCTGCTGCGAGGAAATCTGTC 657
 Db 651 GTTATCAGCTAGCGAGTTAGCTGCTGAGAGAGCTTATTGGAATGG----- 695
 QY 658 AAATTCAGAGAACCTTATCCCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACC 717
 Db 696 -GAAGCAGGGATCTCGTCTCTTCAAGTTCTAGTTTATAATGCAATCCAGCTCAACCAA 754
 QY 718 TTCTTAAGCAATCCAGGAACTACAAATACTAAACAAGCAACACAGCAACACTTAACAG 777
 Db 755 GATTGTCAGAGNACCAATCT-----GACTGCTCACTCAACTTA 794
 QY 778 TCAAGCAAGTCAAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
 Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGGCTTTTACGTGAATGTTGATGTTAAACCTTT 854
 QY 838 GAGTCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCACAATCAACAAGTCG 897

Db 855 ATCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATATCAAGTCTG 914
 QY 898 AACAGCTAGAGGTGTTGCACTGCGACACAGGAGATCATTAACATCTATCCCTTACTCTCA 957
 Db 915 AACCGCAGAGGTGTAGTCTGCTCCTCATGTGTAAACCATTAACCATTTATCCCTTATGAACA 974
 QY 958 AATGTCTGAATTTGAAGAAAGCAATCTGCTATTAATTCCTTCGTTATCGTTATCGTTCAACCA 1017
 Db 975 AATGTCTGAATTTGAAGAAAGCAATCTGCTATTAATTCCTTCGTTATCGTTATCGTTCAACCA 1034
 QY 1018 TTGGGTACCAAGATTCAGGCGCAGAAACCAACCAAGTCCACACCGACTCCGGAAACCTAGTCC 1077
 Db 1035 TTGGGTACCAAGATTCAGAACCAAGAGACCAAGTCCACACCGACTCCAGAACCTAGTCC 1094
 QY 1078 AGGCCGCAACCTGCAACCAATCTTAAATATAGATCAAAATCTTCTTGGTTAGTCAAGT 1137
 Db 1095 AAGTCCGCAACCAAGTCCAGCAATCAATTTGATGGAA-----ATTGGTCAAGAAAGC 1148
 QY 1138 GGTACGAAAGTGTGGGCAAGGATATGTTTCGAAGAAAGGCACTCTCGTTATGTTCT 1197
 Db 1149 TGTTTCAAAAGTAGGCGATGTTTATGTTCTTTGAGGAGATGGAGTTTCTCGTTATATCCC 1208
 QY 1198 TGGCAAGATTTTACCATCTGAACTGTAAATCTTTAAAGCAAGTTATCAAAACAAGA 1257
 Db 1209 AGCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACTGGCCAGCAGGA 1268
 QY 1258 GAGTGTTCACACACTTTAACTGTAAAGAAAGAAATGTTGCTCTCGTGACCAAGATTT 1317
 Db 1269 AAGTTTATCTCATAGCTAGGAATTAAGAAATGACCTCCCATCTAGTATGATCGAATTT 1328
 QY 1318 TTATGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAAGGG 1377
 Db 1329 TTACAATAGGCTTATGACCTTACTAGCAAGATTTCCAGAGATTTACTTGAATATTAAGG 1388
 QY 1378 TCGTAATTTCTGATTTCCAAGCTTTAGACAAATTAATTAGAAGCTTTGAATGATGAATCGAC 1437
 Db 1389 TCGCAAGATTTGATTTTGGAGCTTTGGATAACCTCTGTGAAGCACTCAAGGATGCTCAAG 1448
 QY 1438 TAATAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCATTACCCATCCAGA 1497
 Db 1449 TGATAAAGTCAAGTTAGTGAAGATATTTCTTGGCTCTTTAGTCTCCGATTCGTATCCAGA 1508
 QY 1498 GCGACTTTGGCAACCAATCTCAAAATTTAGTATCTAGTGAAGCAAGTTCTGATTTGTCTCA 1557
 Db 1509 ACGTTTAGGAAACCAATGCGCAATTTACCTACCTGATGATGATGATCAAGTAGCCAA 1568
 QY 1558 ATTAGCTGATAAGTATACAACTGAGTGGTTACATTTTTCATGAAATGATATATATCAG 1617
 Db 1569 GTTGGCAGGCAAGTACACAGCAGAAAGCGGTTATATCTTTGATCCTCGTGATATAACCA 1628
 QY 1618 TGATGAAGGATGATATGTAACGCTCATATGGCCATAGTCACTGATTTGGAAGAAGA 1677
 Db 1629 TGATGAGGGGATGCTTATGTAATCTCAATATGACCCATAGCCACTGATTTAAAGAAGA 1688
 QY 1678 TAGCCTTTCTGATAAGGAAAAAGTTGCAAGCTTCAAGCTTATTAATAAGAAAAAGGTATCCT 1737
 Db 1689 TAGTTTCTGAGCTGAGAGCGGCGAGCCAGCTTATGCTTAAAGAAAGGTTTGAC 1748
 QY 1738 ACCTCCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACTGAGATAGTGCAGCAGCTAT 1797
 Db 1749 CCCTCCCTTCGACAGACCATCAGGATTCAGGAAATCTAGGCAAAAGGAGCAGAGCTAT 1808
 QY 1798 TTACAATCGTGAAGGAAAAACGAATTCCTCGTTCGATTCCTCATATATGTTGA 1857
 Db 1809 CTACAACGAGTGAAGCAGCTAGAGAGTGCCTTGTATGCTATGCTTCAATCTTCA 1868
 QY 1858 GCATACAGTTGAGGTTAAAGCGTAAATTTGATTTATTCCTCATAGGATCATTAACCAFAA 1917
 Db 1869 ATATACGTAGAAGTCAAAAGCGGTAGTTTATCATCACTCATTAATGATGACCATTAACATA 1928
 QY 1918 TATTAAATTTCTGTTTGTATGATCACATACAAAGCTCAATGCTATACCTTGA 1977

Db 1929 CATCAAAATTTAGTGGTTTTCAGCAAGGCTTTATGAGGCACTTAAGGGGTATATCTTTGA 1988
 QY 1978 AGATTTCTTTGCGACGATTAAGTACTACGTAGAACCCCTGACGACGCTCCACATTTCTAA 2037
 Db 1989 GGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGACGCTCCGATTCAGA 2048
 QY 2038 TGATGATGGGGCAATGCCAGTGTGTTAGGCAAGAAAGACCAACAGTGAAGATCC 2097
 Db 2049 TAATGGTTTGGTTAAACGCTAGCGACCATGTTCAAGAAACCAAAATGGTCAAGCTGATAC 2108
 QY 2098 AATAAGAACTTCAAGCGGATGA-----AGAGCCAGT 2121
 Db 2109 CAATCAACGGAAGAAACCAAGCGGAGGAGAAACCTCAGACAGAAAAAACCCTGAGGAAGAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCTCGAAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAACCAACAGGGAACC 2228
 QY 2131 AGAGGAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAGAGTAGAAGCCCA 2190
 Db 2229 AGAAGATCACAGAGGAATCAGAGAAACCTCAGTCTGAGACTGAAAGAGTTGAAGAAA 2288
 QY 2191 ACTCAAGAGCAGAGTGTGCTTGGCAAGTAAACGATTTCTAGTCTGAAAGCCATGC 2250
 Db 2289 ACTGAGAGGCTCAGAGATTTACTTGGAAAAATCCAGGATCCAAATTTCAAGTCCAAATGC 2348
 QY 2251 AACAGAACTCTAGTCTGTTTACGAAATAATTTGACTCTTCAAATTTATGATACAAATAG 2310
 Db 2349 CAAGAGACTCTCAGAGATTTAAATAATTTACTATTTGGCACCAGCAGCAACATAC 2408
 QY 2311 TATCATGGCAGAGCAGAAAAATTTACTTGGCTTTTAAAGGAGTAA 2358
 Db 2409 TATTATGGCAGAGCTGAAAAAATCTATTGGCTTTATTAAAGGAGATTA 2456

RESULT 16

ABX06705
 ID ABX06705 standard; DNA; 2517 BP.

XX AC ABX06705;

XX DT 27-OCT-2003 (revised)

XX DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain coding region #993.

XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 XX ear infection; antinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

XX PD 03-OCT-2002.

XX XX 27-MAR-2002; 2002WO-IB002163.

XX XX 27-MAR-2001; 2001GB-00007658.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Masighani V, Tettelin H, Fraser C;

XX DR WPI; 2003-040579/03.

XX DR P-PSDB; AEU01418.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 XX useful as medicaments for treating or preventing a disease or infection
 XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 XX ear infection.

Claim 6; SEQ ID NO 1985; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2489 identified coding region from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 27-OCT-2003 to standardise OS field)

Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 U; 0 Other;

Query Match	41.4%	Score 990	DB 7	Length 2517	
Best Local Similarity	68.1%	Pred. No. 1.2e-239			
Matches 1478	Conservative	0	Mismatches 636	Indels 57	Gaps 5
QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA-----TAATCGTGTTC	57		
Db	60	TTCCCTATGAACITTGGTCGTCCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC	119		
QY	58	CTATATAGATGGAAAAACAACGCGACGCAAAAAACGGAGAATTTCGACTCCTCATGAGGTTAG	117		
Db	120	TTATATAGATGGTGTGATCAGGCTGGTCAAAAGGCGAANAACCTTGACACAGATGAAGTCAG	179		
QY	118	CAACCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT	177		
Db	180	TAAAGGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT	239		
QY	178	CACCTTCATGGCGACCACTATCATTTATTTACAATGGTAAAGGTTCTTTATGACCGCTATCAT	237		
Db	240	GACCTCTCATGGAGAGCCATTATCATTTACTATAATGCGAAGTGCCCTTATGATGCCATCAT	299		
QY	238	CAGTGAAGATTAICTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAA	297		
Db	300	CAGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTCAGACATTTGCAA	359		
QY	298	TGAGGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAAAATACTATCTGTTTACCTTAAGGA	357		
Db	360	TGAANTCAAGGGTGGTTATGTTTATCAAGGTAGATGGAAAATACTATCTGTTTACCTTAAGGA	419		
QY	358	TGCTGCCCACGCGGATAACGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGACA	417		
Db	420	TGCAGCTCATCGGATAATATTCCGACAAAAGAGAGATTTAAAGCTCAGAGCGAGAACAA	479		
QY	418	TAGTCAACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTGCCTTGGCACGCTTC	477		
Db	480	CAGTCATAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCAGCCAGAC	533		
QY	478	GCAAGGACGCTTACTACAGATGATGGTTATATCTTTTAATGCTTCTGATATCATAGAGGA	537		

QY 1606 TGATATAATCAGTGATGAGGAGATGCATATGTAAAGCCCTCATATGGCCCATAGTCACTG 1665
 Db |||||
 QY 1638 TGATATAACAGTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTG 1697
 Db |||||
 QY 1666 GATTGAAAAGATAGCCCTCTTCTGATAGGAAAAGTTGACGCTCAAGCCTATATAAGA 1725
 Db |||||
 QY 1698 GATTAAAAAGATAGTTGTCTGAAGCTGAGAGGCGGAGCCAGGCTTATGCTAAAGA 1757
 Db |||||
 QY 1726 AAAAGGTATCCTACTCCTCATCTCCAGAGCGGAGATGTTAAAGCAATCCAAGTGGAGATAG 1785
 Db |||||
 QY 1758 GAAAGGTTTACCCCTCTTCGACAGACCATCAGGATTCAGGAATATCTAGGCGAAAGG 1817
 Db |||||
 QY 1786 TGACAGAGCTATTACAAATCGTGTGAAAGGGGAAAACGAATTCACCTCGTTCGACTTCC 1845
 Db |||||
 QY 1818 AGCAGAAGCTATCTACAAACCGCTGAAAGCGAGCTTAAGAGGTGCCACTTGATCGTATGCC 1877
 Db |||||
 QY 1846 ATATATGTTTGAAGATACAGTTGAGTTGAAACCGTAAATTTGATTAATTCCTCATAGA 1905
 Db |||||
 QY 1878 TTACAATCTTCAATATATCTAGAAAGTCAAAAACGGTAGTTTAATCATACCTCATATGA 1937
 Db |||||
 QY 1906 TCATTACCAATATTAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGG 1965
 Db |||||
 QY 1938 CCATTACCAATCAATCAATTTGAGTTGTGACGAAGGCTTTATGAGGACCTTAAGGG 1997
 Db |||||
 QY 1966 CTATACCTTGAAGATTTGTTGGCAGGATTAAGTACTACGTAGAACACACCTGACGAACG 2025
 Db |||||
 QY 1998 GTATACCTTTGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGACG 2057
 Db |||||
 QY 2026 TCCACATCTTAATGATGAGTGGGCAATGCGAGTGAGCATGTGTTAGCGAAGAAGACCA 2085
 Db |||||
 QY 2058 TCCGCAATTCAGATAATGTTTGTGTAACGCTAGCGACCATGTTTCGTAATAATAGGTAGA 2117
 Db |||||
 QY 2086 CAGTGAAGATCCAATTAAGACTTCAAGCGGATGACAGCCAGTACAGGAAGACACCTGC 2145
 Db |||||
 QY 2118 CCAAGACAGTAACCTGATGAAGATAAGAACATGATGAAGTAAGTACGACCACTCACCC 2177
 Db |||||
 QY 2146 TGAGCCAGAG 2156
 Db |||||
 QY 2178 TGATCTGATG 2188
 Db |||||
 RESULT 17
 AAA47602
 ID AAA47602 standard; DNA; 2531 BP.
 AC AAA47602;
 XX 20-OCT-2000 (first entry)
 DT Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.
 DE Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 XX histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX Streptococcus pneumoniae.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2517
 FT /*tag= a
 FT /product= "Sp36D polypeptide"
 XX
 XX WO200037105-A2.
 PN 29-JUN-2000.
 XX
 XX 21-DEC-1999; 99WO-US030390.
 XX
 XX 21-DEC-1998; 98US-0113048P.
 XX
 XX (MEDI-) MEDIMUNE INC.
 PA

Johnson LS, Koenig S, Adamou JE;
 WPI; 2000-452129/39.
 P-PSDB; AAB01466.
 Vaccine useful for prophylaxis and treatment of pneumococcal infections
 such as otitis media, nasopharyngeal and bronchial infections, comprises
 Streptococcus pneumoniae proteins.
 Disclosure; Page 57-58; 70pp; English.
 Although a number of proteins have been suggested as being involved in
 the pathogenicity of Streptococcus pneumoniae, there still remains a need
 to identify polypeptides having epitopes in common from various strains
 of S. pneumoniae in order to utilize such polypeptides in vaccines to
 protect against a wide variety of S. pneumoniae. New vaccine compositions
 are described which comprise a Streptococcus pneumoniae polypeptide (or
 fragments) of 80 - 680 amino acids in length that comprise at least one
 histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 directed against these features. The vaccine is useful in protecting
 against infection by Streptococcus pneumoniae. The vaccine composition
 comprising antibodies to is useful for passive immunization for treating
 CC Pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX
 SQ Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 U; 0 Other;
 Query Match 41.48; Score 990; DB 3; Length 2531;
 Best Local Similarity 68.18; Pred. No. 1.2e-239;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;
 QY 1 TTCTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAGGAAAA---TAATCGTGTTC 57
 Db |||||
 QY 60 TTCTATGAATTTGGTGTCTACCAAGCTGTGAGTTAAGAAAGAGCTAATCAGTTTC 119
 Db |||||
 QY 58 CTATATAGATGGAAGAACAGCGACGAAAGAAACCGAGAAATTTGACTCCTGATGAGGTAG 117
 Db |||||
 QY 120 TTATATAGATGCTGATCAGGCTGTGTCAAAAGGCGAGAAAACTTGACACAGATGAAGTCAG 179
 Db |||||
 QY 118 CAAGCGTGAAGGATCAATGCTGAGCAATCGTCAATCAAGATACAGACCAAGCTATGT 177
 Db |||||
 QY 180 TAAGGGGAGGGATCAACGCCGAAACATCGTCAATCAAGATACGATCAAGTTATGT 239
 Db |||||
 QY 178 CACTTCACATGGGACCACTATCATTTATTCAATGTGTAAGTTCCTTTATGCGTATCAT 237
 Db |||||
 QY 240 GACCTCTCAGGAGACCATTCATTACTATATATGCAAGGTCCCTTATGATGCCATCAT 299
 Db |||||
 QY 238 CAGTGAAGATTTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTAA 297
 Db |||||
 QY 300 CAGTGAAGAGCTCCTCATGAAGATCGGAATTTATCAGTTGAAGGATTCAGACATTTGCA 359
 Db |||||
 QY 298 TGAGGTCAAGGGTGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGA 357
 Db |||||
 QY 360 TGAATCAAGGGTGATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGA 419
 Db |||||
 QY 358 TGCTGCCACGCGATTAACGTCCTGACAAAAGAGGAAATCAATCGACAAAACAAAGACA 417
 Db |||||
 QY 420 TGCAGCTCATCGGATTAATTCGACAAAAGAGAGATTAAACGTGAGAGCAGGAACA 479
 Db |||||
 QY 418 TAGTCAACATCGTGAAGTGAAGTCCAGAAACGATGGTGTCTTGGCCACGTTTC 477
 Db |||||
 QY 480 CAGTCATAATCACGGGGTGTCTCT-----AACGATCAAGCAGTAGTTGCGAGCCAGAGC 533
 Db |||||
 QY 478 GCAAGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
 Db |||||
 QY 534 CCAAGGACGCTATACAGCGATGATGGTTATATCTTCAATGCACTGATATCATGAGGA 593
 Db |||||
 QY 538 TACTGGTGATGCTTATATCGTTCTCATGGAGATCATTTACCATTCATTCCTAAGATGA 597
 Db |||||
 QY 594 CACGGGTGATGCTTATATCGTTCTCAGCGGACCATTAACCATTCCTAAGATGA 653
 Db |||||
 QY 598 GTTATCAGCTAGCAGTTGGCTGCTGACAGAGCCTTCTCTATCTGTCGAGGAATCTGTC 657
 Db |||||

Db 654 GTTATCAGCTAGCGAGTGTAGCTCTCGAAGCCT----- 688
 QY 658 AATTCAGAACTTATCGCGGCAAAATAGCGATATACTTCAAGAACAACTGGGTACC 717
 Db 689 -----ATTGGAATGGGAAGCAGGATCTCGCTCTCTTCAAGTTCTAGTTATATATGC 740
 QY 718 TTCTGTAAGCAATCCAGGAACATAAATACTAACAGCAACAAAGCAAGCACTAACAG 777
 Db 741 AATTCAGCTCAACCAAGATTGTGAGAGAACCAATCTGACTGTCTCTCAACTTATCA 800
 QY 778 TCAAGCAAGTCAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTCAAACTGCTTT 837
 Db 801 TCA---AATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTTAAACCCCT 857
 QY 838 GAGTCAACACATGTAGATCTGATGGCCTTGTCTTTGATCGAGCAAAATCAAGTCG 897
 Db 858 ATCAGAACCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAAATCAAGTCG 917
 QY 898 AACAGCTAGAGGTGTTCAGTGCACACGAGATCAATACCACTTCACTCCCTTACTCTCA 957
 Db 918 AACCCGACAGGTGTAGCTGTCCCTCATGTGAACATTTACCCTTTATCCCTTATGAACA 977
 QY 958 AATGCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTCTTCTGTTTCAAAACCA 1017
 Db 978 AATGCTGAATTTGGAAGAACGAATTCGCTCGTATTTATTCCTCTTCTGTTTCAAAACCA 1037
 QY 1018 TTGGGTACAGATTCAGGCGCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
 Db 1038 TTGGGTACAGATTCAGGCGCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1097
 QY 1078 AGCCCGCAACCTGCACCAAACTCTTAAATAGACTCAAAATTTCTC-----TTT 1125
 Db 1098 AAGTCGCAACCTGCACCAAACTCTTAAATAGACTCAAAATTTCTC-----TTT 1157
 QY 1126 GGTATGCTAGCTGTGGAAGAACGAATTTGGGAAGGATATGATTGGAAGAAAGGCAATCTC 1185
 Db 1158 GGTCAAGAAAGCTGTTCGAAAGTAGCGGATGTTATGTTCTTTGAGGAGAAATGGAGTTTC 1217
 QY 1186 TCGTTATGCTTTGCAAGAGATTACCACTGCAACTGTAATCTGTAATCTTGAAGCAAGTT 1245
 Db 1218 TCGTTATATCCGACCAAGATCTTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACT 1277
 QY 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATTTGCTCTCTCG 1305
 Db 1278 GGCCAGCAGGAAAGTTTATCTCATAGCTAGAGCTAGAAAACCTGACTCCCATCTAG 1337
 QY 1306 TGACCAAGAAATTTATGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGT 1365
 Db 1338 TGATCGAGAATTTTACAATAAGCCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
 QY 1366 TGNAAATAGGTCGTAAATCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAA 1425
 Db 1398 TGATATAAGGTCGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAACGACTCAA 1457
 QY 1426 TGATGAATCGACTAATAAAGAAAAATTTGATGATGATTTTATGGCAATCTAGCACCAT 1485
 Db 1458 GGATGTCCTCAAGTAAAGTCAAGTTAGTGATGATATCTTGCTCTTCTAGTCCGAT 1517
 QY 1486 TACCATTCAGAGGACTTGGCAACCAAAATCTCAAATTTAGATATCTGAAAGCAAGT 1545
 Db 1518 TCGTTCATCCAGAACGTTTAGGAAAAACCAATGCGCAATTTACCTACACTGATGAGAT 1577
 QY 1546 TCGTATTTCTAATAGCTGATAGTATACAGCTCAGATGTTTATCTTTTATGATGAACA 1605
 Db 1578 TCAAGTAGCAAGTTGGCAGGCAAGTATACACAGAAAGCGGTTATATCTTTGATCCTCG 1637
 QY 1606 TGATATAATCAGTATGAAAGGAGATGCATATGTAAGCCTCATATGGGCATAGTCACTG 1665
 Db 1638 TGATATAACAGTATGAGGGGATGCTTATGTAATCTCAATATGACCCATAGCCACTG 1697
 QY 1666 GATTGGAAGATAGCCTTTCTGATAAGGAAAAAGTTGAGCTCAAGCCTTACTATAAGA 1725

Db 1698 GATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATCTTAAGA 1757
 QY 1726 AAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACCTGAGATAG 1785
 Db 1758 GAAAGGTTTGAACCCCTCTTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAG 1817
 QY 1786 TGACGAGCTATTTTCAATCGTGTGAAGGGGAAAAACGAATTCCTACTGTTCCGACTTCC 1845
 Db 1818 AGCAGAAAGCTATCTCAACCGGTGAAGAGCGCTAAGAAAGGTGCCACTTGTATCGTATGC 1877
 QY 1846 ATATATGTTGAGCATACATAGTTGAGGTTAAAAACGGTAATTTGATTTATCTCTCATAGGA 1905
 Db 1878 TTACATCTTCAATATCTAGTAGAGTCAAAACGGTAGTTTAACTCATCTCATATGA 1937
 QY 1906 TCATTAACCAATATTAATTAATTTGCTTTGTTGTTGATGATCAACATAAAAGCTCCAAATGG 1965
 Db 1938 CCATTAACCAATATTAATTAATTTGAGTGGTTTTCAGGAAGGCTTTTATGAGGCACTAAGGG 1997
 QY 1966 CTATACCTTTGGAAGATTTGTTTTCGACGATTAAGTACTAGTAGACACCTGACCAACG 2025
 Db 1998 GTATATCTTGAAGATCTTTTTCGACGACTGTCAAGTACTATGTCGAACATCCAAACGACG 2057
 QY 2026 TCCACATCTTATGATGATGGGCAATGCCATGAGCTAGCTAGTGTAGGCAAGAAAGACCA 2085
 Db 2058 TCGCATTCAGATAATGTTTGGTAAACCTAGCGACCATGTTTCGTAATAAATAGGTAGA 2117
 QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGACCGAGTAGAGGAAACACCTGC 2145
 Db 2118 CCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAAGTGAAGCACTCACCC 2177
 QY 2146 TGAGCCAGAAG 2156
 Db 2178 TGAATCTGATG 2188

RESULT 18

ABS56454_08

Continuation (9 of 22) of ABS56454 from base 800001 (Streptococcus pneumoniae type 4 str
 WP Sequence split into 22 fragments LOCUS ABS56454 Accession Abs56454

WP	Fragment Name	Begin	End
WP	ABS56454_00	1	110000
WP	ABS56454_01	100001	210000
WP	ABS56454_02	200001	310000
WP	ABS56454_03	300001	410000
WP	ABS56454_04	400001	510000
WP	ABS56454_05	500001	610000
WP	ABS56454_06	600001	710000
WP	ABS56454_07	700001	810000
WP	ABS56454_08	800001	910000
WP	ABS56454_09	900001	1010000
WP	ABS56454_10	1000001	1110000
WP	ABS56454_11	1100001	1210000
WP	ABS56454_12	1200001	1310000
WP	ABS56454_13	1300001	1410000
WP	ABS56454_14	1400001	1510000
WP	ABS56454_15	1500001	1610000
WP	ABS56454_16	1600001	1710000
WP	ABS56454_17	1700001	1810000
WP	ABS56454_18	1800001	1910000
WP	ABS56454_19	1900001	2010000
WP	ABS56454_20	2000001	2110000
WP	ABS56454_21	2100001	2162598

Query Match

Best Local Similarity 41.4%; Score 990; DB 7; Length 110000;
 Pred. No. 6.3e-239;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 1 TTCTTACAGTTGGGACTGTATCAAGCTAGAACGCTTAAGGAAA---TAATCGTGTTC 57
 Db 38008 TTCTTATGAATTTGTCGTCTCAAGCTGTGTCAGGTTAAGAAAGAGTCTAATTCGAGTTTC 38067
 QY 58 CTATATAGATGMAAACAGCGACCAAAAAACCGAGAAATTTGACTCTCTGATGAGTTAG 117

```
Db 38068 TTATATAGATGGTGATCAGCGTGGTCAAAAGGCGAGAAAACCTTCACACCAAGATCAAGTCAG 38127
Qy 118 TCCTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTGTTAAATACTTTGAAAGCAAGTT 1245
Db 39166 TCGTTATATCCAGCCAGGATCTTTTCAGCAGAAACAGCAGCAGGCGATTCATAGCAAACT 39225
Qy 1246 ATCAAAACAAGAGAGTGTTTTCACACACTTTAACTGCTTAAAAAAGAAAATGTTGCTCTCG 1305
Db 39226 GGCACAGCAGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAAACCTGACCTCCCATCTAG 39285
Qy 1306 TGACCAAGATTTTATGATAAAGCATATACTCTGTTAACTGAGCTCATAAAGCCTTGTT 1365
Db 39286 TGATCGAGAAATTTTACAAATAGGCTTAAGACTTACTAGCAGAAATTCACAGATTTACT 39345
Qy 1366 TGNAAATAAGCGGTGTAATTTCTGATTTTCCAAGCCTTAGACAAATTTATTAGAACCTTTGAA 1425
Db 39346 TGATATAAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAAACCCTGTTGGAACGACTCAA 39405
Qy 1426 TGATGAATTCGACTAATAAAGAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAAT 1485
Db 39406 GGAATGTCCCAAGTGATAAAGTCAAGTTAGTGGATGATTTCTTGCTTCTTAGCTCCGAT 39465
Qy 1486 TACCATCCAGAGCGACTTCGCAAAACCAATTTCTCAAAATGAGTATATCTGAAGACGAGT 1545
Db 39466 TCGTCATCCAGAGCTTTGAGAAAACCAATTCGCCAAATTTACCTACTGATGATGAT 39525
Qy 1546 TCGTATTGCTCAATTAGCTGATTAAGTATACAAAGCTCAGATGGTTTACATTTTGTATGAACA 1605
Db 39526 TCAAGTAGCCAAAGTTGCGAGGCAAGTACACACAGAAAGCGGTTATATCTTTGATCCTCG 39585
Qy 1606 TGATAAATCAGTGATGAAGAGATGATATATGTAACCCCTCATATGGGCCATAGTCACTG 1665
Db 39586 TGATATAACCAAGTATGAGGGGATGCTATGTAACCTCCACATATGACCCATAGCCACTG 39645
Qy 1666 GATTGGAAGAGATAGCTTCTGATAAGGAAAAGTTGCAAGCTCAAGCTCATATAAAGA 1725
Db 39646 GATTAAAAAGATAGTTTGTCTGAGCTGAGAGCGCGAGCCAGGCTTATGCTAAGA 39705
Qy 1726 AAAGGTATCTACTCTCCATCTCCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAG 1785
Db 39706 GAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGG 39765
Qy 1786 TGCAGAGCTATTATCAATCGTGTGAAGGGGAAAACGAAATTCACCTCGTTCGACTTCC 1845
Db 39766 AGCAGAGCTATCTTACAAACCGCGTGAAGCAGCTTAAGAGGTGCCACTTGATCGTATGCC 39825
Qy 1846 ATATATGTTGAGCATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAAGGA 1905
Db 39826 TTACAATCTCAATATCTAGTAGAGTCAAAAACGGTAGITTTATCATACCTCATATGA 39885
Qy 1906 TCATTACCAATAATTAATTTGCTTGGTTTGTATGATCAACATACAAAGCTCCAAATGG 1965
Db 39886 CCATTACCAATAACATCAAAATTTGAGTGTGTTGACGAAGGCGCTTTATGAGGCACTAAGG 39945
Qy 1966 CTATACCTTGGAAGATTTGTTGCGAGATTTAGTACTAGCTAGAACACCCCTGACGACG 2025
Db 39946 GTATACCTTGAGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGACG 40005
Qy 2026 TCCACATTTCAATGATGGATGGGCAATGCGCAGTGCAGTGTGTAGGCAAGAAGACCA 2085
Db 40006 TCCGCAITTCAGATAATGTTTGTGTAACGCTAGCGACCATGTTCTGTAAAAATAAGGTAGA 40065
Qy 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGAGCCAGTAGGAGAAACACCTGC 2145
Db 40066 CCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAAGTGAAGCAACTCACCC 40125
Qy 2146 TGAGCCAGAG 2156
Db 40126 TGAATCTGATG 40136
Qy 38068 TTATATAGATGGTGATCAGCGTGGTCAAAAGGCGAGAAAACCTTCACACCAAGATCAAGTCAG 38127
Qy 118 CAAGCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATTAACAGAACGAGCTATGT 177
Db 38128 TAAGAGGAGGGGATCAACCCGAAACAAATCGTCATCAAGATTAACGAGTCAAGGTTATGT 38187
Qy 178 CACTTCAATGGGACCACTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 237
Db 38188 GACTCTCATGGAGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 38247
Qy 238 CAGTGAAGAATTAATCTATGAAGAATCCAAATTAAGCTAAAAAGATGAGGATATGTTAA 297
Db 38248 CAGTGAAGAGCTCTCATGAAGATCCGAATTAATCACTGAAGGATTCAGACATTTGTCAA 38307
Qy 298 TGAGGTCAAGGTTGGATATGTTATCAAGTATGATGGAATATCACTGTTTACCTTAAAGGA 357
Db 38308 TGAATCAAGGTTGGTATGTTATCAAGGTATGATGGAATATCACTGTTTACCTTAAAGGA 38367
Qy 358 TGCTGCCACGCGGATTAACGTCCTGACAAAAGAGGAATCAATCGACAAAAACAAGAGCA 417
Db 38368 TGCAGCTCATGGGATTAATTCGACAAAAGAGAGATTAAGCTCAGAGCAGGAACA 38427
Qy 418 TAGTCAACATCGTGAAGGTGGAATCCCAAGAAAGTGGTGTCTGTCCTTGGCAGCGTTC 477
Db 38428 CAGTCAATAACACGGGGTGGTCTCT-----AACGATCAAGCAGTAGTTGAGCCAGAGC 38481
Qy 478 GCAAGGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db 38482 CCAAGGACGCTATCAACCGGATGATGTTATCTTCAATGCACTCTGATATCATAGAGGA 38541
Qy 538 TACTGTGTATGCTTATATCGTTCTCATGGAGATCATTAACCATTAATTCCTAAGAATGA 597
Db 38542 CACGGTGTATGCTTATATCGTTCTCTCGCGGACCAATTAACCATTAATTCCTAAGAATGA 38601
Qy 598 GTTATACGCTAGCAGTTGGTGTCTGCAAGAGCCTTCTATCTGGTCGAGAAATCTGTC 657
Db 38602 GTTATACGCTAGCAGTTAGTGTCTGCAAGAGCCT----- 38636
Qy 658 AAATTCAGAACCTATCGCCGACAAAATAGGATAACACTTCAAGAACAAACTGGGTACC 717
Db 38637 -----ATTGGAATGGAGAGGAGGATCTGCTCTTCAAGTTCTAGTTAATATGC 38688
Qy 718 TTCTGTGAAGCAATCCAGGAATCAAAATPACTAACAAGCAACAAACAGCAACACTAAG 777
Db 38689 AAATCCAGCTCAACCAAGATGTCAGAGAACCAATCTGACTCTCACTCCAACTTATCA 38748
Qy 778 TCAAGCAAGTCAAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
Db 38749 TCA---AAATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCT 38805
Qy 838 GAGTCAACGATGTAGAACTGTGATGGCTTGTCTTTGATCCAGACAAATCAACAAGTCG 897
Db 38806 ATCAGAACGCAATGTGGAATCTGATGGCTTATTTTCAGCCAGCGCAATCAACAAGTCG 38865
Qy 898 AACAGCTAGAGGTGTGCAAGTCCACACGAGATCATTAACACTTCACTTCTTACTCTCA 957
Db 38866 AACGCCAGAGGTGTAGCTGTCCCTCATGTGTAACCAATTAACCACTTATTCCTTATGAACA 38925
Qy 958 AATGCTGAATTTGNAAGAACAAATCGCTGCTTATTTCCCTTCTGTTATCGTTTCAACCA 1017
Db 38926 AATGCTGAATTTGNAAGAACAAATCGCTGCTTATTTCCCTTCTGTTATCGTTTCAACCA 38985
Qy 1018 TTGGGTACCAAGATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTTAGTCC 1077
Db 38986 TTGGGTACCAAGATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTTAGTCC 39045
Qy 1078 AGGCCCGCAACTGCAACCAATCTTAAATATAGATCAAAATCTTCT-----TTT 1125
Db 39046 AAGTCGCAACTGCAACCAATCTTAAATATAGATCAAAATCTTCT-----TTT 1125
Qy 1126 GGTAGTCAGCTGGTACCAAGATTTGGGAGAGATGATGTTTTCGAAGAAAAGGCGATCTC 1185
Db 39106 GGTCAAGAAAGCTGTTCGAAAAGTAGGCGATGGTTATGTTCTTTGAGGAGATGGAGTTTC 39165
```

RESULT 19
AAV27356
ID AAV27356 standard; DNA; 2290 BP.

XX	AAV27356;	
AC	02-OCT-1998 (first entry)	
XX	Streptococcus pneumoniae SP0042 nucleotide.	
XX	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;	
XX	detection; pneumonia; otitis media; meningitis; ss.	
XX	Streptococcus pneumoniae.	
OS		
XX		
PH	Key Location/Qualifiers	
FT	CDS 2..2290	
FT	/*tag= a	
FT	/product= "SP0042"	
FT	/transl_except= (pos:152..154,aa:Xaa)	
FT	/transl_except= (pos:1406..1408,aa:Xaa)	
FT	/transl_except= (pos:1430..1432,aa:Xaa)	
FT	/note= "no stop codon given; Xaa is unspecified"	
XX		
PN	W09818930-A2.	
XX		
XX	07-MAY-1998.	
XX	30-OCT-1997; 97WO-US019422.	
PF		
XX	31-OCT-1996; 96US-0029960P.	
PR		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX	Kunsch CA, Choi GH, Johnson LS, Hromockyj A;	
PI		
XX	WPI; 1998-272224/24.	
DR	P-PSDB; AAW55095.	
XX		
DR	Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae	
PT	- or their epitope-containing fragments, useful in protective or	
PT	therapeutic vaccines, and for diagnosis.	
XX		
PS	Claim 1; Page 61-62; 118pp; English.	
XX		
CC	The present sequence encodes a protein from Streptococcus pneumoniae. The	
CC	nucleic acid sequence encoding the Streptococcus pneumoniae protein can	
CC	be useful in vaccines for inducing protective antibodies against	
CC	Streptococcus pneumoniae, for treatment or prevention of infection e.g.	
CC	pneumonia, otitis media or meningitis. Probes based on the nucleic acid	
CC	are used to detect Streptococcus infection (by usual hybridisation or	
CC	amplification methods), also for isolating Streptococcus genes or their	
CC	allelic variants. The protein can be used similarly to detect specific	
CC	antibodies in standard immunoassays, especially for diagnosing or	
CC	monitoring infections. Antibodies which bind the protein are used to	
CC	detect corresponding antigens, to purify the protein and for passive	
CC	immunisation (optionally coupled to a toxin). Vaccines are administered,	
CC	e.g. by injection, orally or through the skin, typically at 0.01-1000	
CC	(especially 10-300) mu g/ml per dose	
XX		
SQ	Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;	
	Query Match 41.3%; Score 987.6; DB 2; Length 2290;	
	Best Local Similarity 67.7%; Pred. No. 4.8e-239;	
	Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;	
QY	1 TTCTTACGAGTTGGAGCTGTATCAAGCTTAGAACGTTTAAGGAAA--TATCGTGTTC 57	
Db	4 TTCTATGAAGTTGGTGTGTGTACCAAGCTGGTTCAGGTTAAGAGAGTCTAATCGAGTTTC 63	
QY	58 CTATATAGATGGAAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTTAG 117	
Db	64 TTATATAGATGGTGTATCAGGCTGGTCAAAGGCGAGAAACTTGACACCAAGATGAAGTCAG 123	
QY	118 CRAAGCTGAGGAATCAATGCTGTAGCAGAAATCGTCAATCAAGATACACAGCAAGGCTATGT 177	

QY 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAAATCTGCTCCTCG 1305
 Db 1219 GGCCAGCAGAGAAAGTTATCTCATAGCTAGGAGCTAAGAAAACACTGACCTCCCATCTAG 1278
 QY 1306 TGACCAAGAAATTTATGATTAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
 Db 1279 TGATCGAAGATTTTACATTAAGCTTTATGACTTACTAGCAAGATTCACCAAGATTACT 1338
 QY 1366 TGNAAATPAAGGGTGTAAATCTGATTTTCAAGCCTTAGACAAATATTAGAACCTTTGAA 1425
 Db 1339 TGATAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTGTGAACGACTCAA 1398
 QY 1426 TGATGATCGACTTAATAAAGAAAATTTGGTAGATGATTTATGTCATTCTTAGCACCAAT 1485
 Db 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATTCTTGCCCTCTTAGTCCCGAT 1458
 QY 1486 TACCCATCCAGAGCGACTTTGGCAACCAAAATTTCTCAAATTTGAGTATACTGAAGACGAAGT 1545
 Db 1459 TCGTCATCCAGAACGTTTAGGANAACCAATGCGCAAAATTTACCTACACTGATGATGAT 1518
 QY 1546 TCGTATTTGCTCAATAGCTGATAGTATACAACTGATGATGATGATGATGATGATGATGAT 1605
 Db 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACGAGACGGTTTATCTTTGATCTCTCG 1578
 QY 1606 TGATATAATCAGTCAATGAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTG 1665
 Db 1579 TGATATAACAGTATGAGGGGATGCTTGTAGTCACTCCCATATGACCCATAGCCACTG 1638
 QY 1666 GATTGGAAGAGATAGCCTTTCTGTAAGGAAAAAGTTGACGCTCAAGCCTATACATAAGA 1725
 Db 1639 GATTAAAAAGATAGTTGCTGAGCTGAGAGAGCGGACCCAGGCTTATGCTAAAGA 1698
 QY 1726 AAAAGGTATCCTACTCCTCAGAGCGAGATGTTAAAGCAATTCACACTGAGATAG 1785
 Db 1699 GAAAGGTTTGAACCTCCTTCCAGACAGACCATCAGGATTCAGGAAATATCTGAGGCAAAAGG 1758
 QY 1786 TGCAGAGCTATTTACATCTGTTGAAAGGGGAAAAACGAATTCACCTCTGCTGACTTCC 1845
 Db 1759 AGCAGAGCTATCTACAAACCGGTGAAAGCAGCTTAAGAGGTGCACTTGATGATGATGCC 1818
 QY 1846 ATATATGTTGAGCATACAGTTGAGGTTTAAACCGTAAATTTGATTTCTCTCATAGA 1905
 Db 1819 TTACAATCTTCAATATACTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCATTA 1878
 QY 1906 TCATTACATTAATTAATTTGCTTGTGTTGATGATCAACATACAAAGCTCCAAATGG 1965
 Db 1879 CCATTACCATTAACATCAAAATTTGAGTGGTTTGAAGAGGCTTTTATGAGGCACTTAAGGG 1938
 QY 1966 CTATACCTTGAAGATTTGTTTGGACGATTAAGTACTAGTACGTAACACACCTGACGAACG 2025
 Db 1939 GTATACCTTGAAGATCTTTTGGACGCTGTCAGTACTATGTCGACATCAACAGCAAG 1998
 QY 2026 TCCATTTCTATGATGATGGGCAATGCCAGTGAGATGTTGTTAGGCAAGAAAGACCA 2085
 Db 1999 TCCGATTCAGATATGTTTGGTGAACGCTAGCAGCCTGTTCAAGAAACAAAAATGG 2058
 QY 2086 CAGTGAAGATCCAATAGAACTTCAAGCGGATGAAGAGCAGTAGAGGAAACACCTGCG 2145
 Db 2059 TCAAGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118
 QY 2146 TGAGCCGAAGTCCCTCAAGTAGAGA 2171
 Db 2119 TGAGGAAGAAACCTTCGAGAGAGA 2144

RESULT 20

ABQ84824

ID ABQ84824 standard; DNA; 2290 BP.

XX

AC ABQ84824;

XX

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; gene; ds.
 KW Streptococcus pneumoniae.
 OS US2002061545-A1.
 XX 23-MAY-2002.
 XX 22-JAN-2001; 2001US-00765272.
 XX 30-OCT-1997; 97US-00961083.
 XX (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX Choi GH, Kunsch CA, Barash SC, Dillon PU, Dougherty B, Fannon MR;
 PI Rosen CA;
 PI WPI; 2002-479261/51.
 DR P-PSDB; ABP54589.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX Claim 1; Page 28-29; 70pp; English.
 PS ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;
 SQ
 Query Match 41.3%; Score 987.6; DB 6; Length 2290;
 Best Local Similarity 67.7%; Pred. No. 4.8e-239;
 Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
 QY 1 TTCTTACAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAA--TAATCGTGTTC 57
 Db 4 TTCTTATGAACCTTGGTCTCACCAGCTGGTTCAGGTAAAGAAAGAGTCTAATCGAGTTTC 63
 QY 58 CTATATAGATCGAAAAAAGCGAGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGTTAG 117
 Db 64 TTATATAGATGCTATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACAGATGAAGTCAG 123
 QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAATCGTATCAAGATAACAGACCAAGGCTATGT 177
 Db 124 TAAGAGGGAGGGGATCAACGCCGAACAAATNGTATCAAGATTACGGATCAAGGTTATGT 183
 QY 178 CACTTCAATCGCGACCACTATCATTTATCAATGGTAAAGTTCCTTATGACGCTATCAT 237
 Db 184 GACCTCTCATGAGACCACTTATCTACTAATAGGCAAGGTTCTCTTATGATGATCAT 243
 QY 238 CAGTGAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
 Db 244 CAGTGAAGAGCTCTCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTTGCAA 303
 QY 298 TGAGTCAAGGTTGATATGTTATCAAGGTAGATGTTATCAAGGTAGTAAATACTATGTTTACCTTAAGGA 357


```
Db 911 TCA--AAATCAAGGGGAARACATTTCAAGCCTTTACGGTAATGTATGCTAAACCCCTT 967
QY 838 GAGTCAACGACATGTAGAAATCTGATGGCCCTTGCTTTGATGCCAGACAAATCAAGTCG 897
Db 968 ATCAGAACCGCATGTAGAAATCTGATGGCCCTTATTTTCGACCCAGCGCAAAATCAAGTCG 1027
QY 898 AACAGCTAGAGGTGTTGCAAGTCCACACGAGAGATCAATACCACTTCATCCCTTACTCTCA 957
Db 1028 AACCGCCAGAGGTGATGCTCCCTCATGTATACCAATTAACCTTTATCCCTTATGAAACA 1087
QY 958 AATGTCTGAATTTGAAGAAACGAATCGCTCGTATTAATTCCTCTGTTTATCGTTCAAAACCA 1017
Db 1088 AATGTCTGAATTTGAAGAAACGAATTCGTCTATTAATTCCTCTGTTTATCGTTCAAAACCA 1147
QY 1018 TTGGGTACAGATTCAGGCCAGAAACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db 1148 TTGGGTACAGATTCAGAACAGAAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCC 1207
QY 1078 AGGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
Db 1208 AAGTCTGCACTTCGACCAAAATCTCAACGAGTCCAGCAATCCAATTCATGAGAAAT 1267
QY 1126 GGTGAGTACGTGTCGCAAAAGTTGGGGAAGGATATGTATTCGAGAAAGAGGCATCTC 1185
Db 1268 GGTCAAGAAAGCTGTTCCGAAAGTAGGCGATGGTTATGTCTTTGAGGAGAAATGGAGTTTC 1327
QY 1186 TCGTTATGCTTTTCGCAAAAGATTTACCATCTGAACATGTTTAAATCTTGAAGCAAGTT 1245
Db 1328 TCGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACT 1387
QY 1246 ATCAAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAATGTTGCTCCTCG 1305
Db 1388 GGCCACAGCAAGATTTTATCTCATAGTAGTAGGAGCTAAGAAACTGACCTCCCATCTAG 1447
QY 1306 TGACCAAGAAATTTATGAATAAGCATAPATCTGTTAACTGAGGCTCATAAAGCCTTGTT 1365
Db 1448 TGATCGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1507
QY 1366 TGNAANTAAGGTCGTAATCTGTAATTTCCAGGCTTAGACAAATTTATTAGAAGCTTGAA 1425
Db 1508 TGATAATAAGGTCGCAAGATTTGAGGTTTGGATAACCTGTTGGAACGACTCAA 1567
QY 1426 TGATGAATCCACTAATAAGAAAAAATTCGTAGATGATTTATTTGGCAATTCCTAGCACCAAT 1485
Db 1568 GGATGCTCAAGTGATAAGTCAAGTTAGTGGATGATATCTTGCTTCTTAGTCCGAT 1627
QY 1486 TACCCATCCAGAGCGACTTGGCAACCAAAATCTCAAAATTGAGTATACCTGAGACGAAGT 1545
Db 1628 TCGTCAATCCAGACGTTTAGGAAACCAAAATGCGCAAAATTACCTACACTGATGATGAGAT 1687
QY 1546 TCGTATTGCTCAATTAGCTGATAAGTATACAACTCAGATGGTTACATTTTGTATGAACA 1605
Db 1688 TCAAGTAGCAAGTTGGCAGCAAGTACACAAAGAGACGGTTATATCTTTGATCTCTCG 1747
QY 1606 TGATATAATCAGTCATGAAGAGAGATGTCATATGTAAACGCTCATATGGGCCATAGTCACTG 1665
Db 1748 TGATATAACAGTCATGAGGGGGATGCTATGTAACTCCACATATGACCCATAGCCACTG 1807
QY 1666 GATTGGAAGATAGCCCTTCTGATAGGAAAAAGTTGCAGCTCAAGCCTATATAAGA 1725
Db 1808 GATTAAAAAAGATAGTTGCTGTAAGCTGAGAGAGCGGACCCAGGCTTATGCTAAGA 1867
QY 1726 AAAAGGTATCTCACTCCATCTCCAGACGCGAGATGTTAAAGCAATCCAACTGAGAGATAG 1785
Db 1868 GAAAGGTTTGACCCCTCTCTCGACAGACCACGAGATTCAGAAATTAAGGCAAAAGG 1927
QY 1786 TGCAGAGCTATTTACAAATCTGTGAAAGGGGAAAAAGCAATTCACCTGTTCCACTTCC 1845
Db 1928 AGCAGAAAGCTATCTACAAACCGCGTGAAGAGAGCTAAGAAAGGTGCACTTGCATGATGCC 1987
QY 1846 ATATATGTTGAGCATACAGTTGAGGTTTAAAGCGTAAATTTGATTTATCTCTCATAGA 1905
```

```
Db 1988 TTACAATCTTTCAATATATCTAGTAAGTCAAAAAACGGTAGTTTAAATCATCATCTCATTATGA 2047
QY 1906 TCATTACCAATAATTAATTTTCCTTGGTTTGTATGATCAACATACAAAGCTCCAAATGG 1965
Db 2048 CCATTACCAATAATTAATTTTGTAGTGGTTTGTAGAGGCTTTATGAGGCACTTAAGG 2107
QY 1966 CTATACCTTTGGAAGATTTGTTTTCGAGAGTAAAGTACTAGTACAGAACCCCTGACGAACG 2025
Db 2108 GTATAGCTTTGAGATCTTTTTCGGACTGTCAAGTACTATGTGCAACATCCAAACGAACG 2167
QY 2026 TCACATTTTAATGATGGATGGGCAATGCCAGTGGACATGTTTAGGCAAGAAAGACCA 2085
Db 2168 TCGCATTCAGATAATGTTTTCGTACGCTAGTACCATGTTTCGTAAAAAATAAGGCAGA 2227
QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGACCCAGTAGAGGAAACACCTGC 2145
Db 2228 CCAAGATAGTAACCTGATGAAGATTAAGAAACATGATGAAGTAAGTGAGCCACTCACCC 2287
QY 2146 TGAGCCAGAAAG 2156
Db 2288 TGAATCTGATG 2298

RESULT 23
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
AC ABK15104;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11-2.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 114..2630
FT /tag= a
FT /product= "BVH-11-2"
FT /note= "The gene is flanked by sequences from the vector
SP64, no information on which is given in the
specification"
XX
BN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA000908.
XX
PR 20-JUN-2000; 2000US-0212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX P-PSDB; AAU75934.
XX
DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteraemia.
XX
PS Example 3; Fig 5; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
```

CC individual susceptible to these disorders. (iii) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, Group A
 CC Streptococcus such as Streptococcus pyogenes, Group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae, or
 CC Streptococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (iii) encoding (i) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (iii) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
 CC peptides described in the method of the invention
 XX
 SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
 Query Match 41.1%; Score 980.8; DB 6; Length 2639;
 Best Local Similarity 67.8%; Pred. No. 2.6e-237;
 Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;
 QY 1 TTCTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTC 57
 DB 173 TTCTATGAACCTTGTGTCGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 232
 QY 58 CTATATAGATGAAACACAGCGACGCAAAAACGGAGAAATTTGACTTCCTGATGAGTTAG 117
 DB 233 TTATATAGATGTTGATCAGCGTGGTCAAAAGGAGGAGAAATTTGACACCGAGTGAAGTCAG 292
 QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGT 177
 DB 293 TAAGAGAGAGGAGTCAACGCCGACCAAAATTTGATCAAGATTAACGATCAAGTTATGT 352
 QY 178 CACTTCAATGCGGACCACTATCATTAATTAATGATGTAAGGTTTCCTTATGAGCTATCAT 237
 DB 353 GACCTCTCATGAGACCAATTAATTAATGATGTAAGGTTTCCTTATGATGCCATCAT 412
 QY 238 CAGTGNAGATTAATCAATGAGATCAAACTATAAGCTTAAAGATGAGGATATGTTAA 297
 DB 413 CAGTGAAGAACTTCTCATGAAGATCCGAAATTAATCAATGATGTAAGGATTCAGACATTTGTC 472
 QY 298 TGAGGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATTAATGATTAAGGATTAAGGA 357
 DB 473 TGAATCAAGGTTGGATATGTTATCAAGGTAGATGAAATTAATGATTAAGGATTAAGGA 532
 QY 358 TCGTCCCAACGGGATAACGTCGTAACAAAGAGGAAATCAATTCGACAAACCAAGAGCA 417
 DB 533 TCGCGCCCATGCGGACCAATTAATTCGACAAAGAGAGATTAACAGTTCAGAGCAGGAAACA 592
 QY 418 TAGTCAACATCGTGAAGTGGAACTCCAGAAACGATGCTGTTGCTTGGCAGGTTTC 477
 DB 593 CAGTCAATATCAT-----AACTCAAGACAGATAATGCTGTTGTCAGCCAGAGC 643
 QY 478 GCAAGGACGCTTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
 DB 644 CCAAGGACGTTATACACCGATGATGAGTATATCTTCAATGATCTGATATCATTTAGGA 703
 QY 538 TACTGTGATGCTTATATGTTTCTCATCGAGATCAATTAACATTAATTCCTTAAGATGA 597
 DB 704 CACGGTGTATGCTTATATGTTTCTCATCGGACCAATTAACATTAATTCCTTAAGATGA 763
 QY 598 GTTATCAGCTAGCGAGTTCGCTGTCAGAGCGCTTCTTATCTGTTGTCGAGGAATCTGTC 657
 DB 764 GTTATCAGCTAGCGAGTTCGCTGTCAGAGCGCTTCTTATCTGTTGTCGAGGAATCTGTC 798
 QY 658 AAATTCAGAACTTATCGCGGACAAAAATAGCGATAACATTCAGAAACAACTGGGTACC 717
 DB 799 -----ATTGGAATGGGAGCAGGGATCTCGTCTCTTTCAAGTTCTTAGTTATATGC 850
 QY 718 TTCTGTGAAGCAATCCAGGAATCAAAATTAACACAGCAACCAACAGCAACTAACAG 777
 DB 851 AAATCCAGTTCAACCAAGATTGTTCAGAGAACCAACATCTGACTGTCACTCCAACCTTATCA 910

QY 778 TCAAGCAAGTCAAGATATGACATTTGATAGTCTCTTTGAAAACAGCTCTTACAAAACGCTTT 837
 DB 911 TCA---AAATCAAGGGGAAAACATTTTCAAGCCCTTTTACGTGAATTTGATGCTAAAACCTT 967
 QY 838 GAGTCAACACATGTAGATATCTGATGGCTTGTCTTTGATCCAGACAAATCAAGTGTG 897
 DB 968 ATCAGAACCCATGTAGATATCTGATGGCTTTATTTTCGACCCAGCGCAATCAAGTGTG 1027
 QY 898 AACAGCTAGAGTGTGTCAGTCCACACGAGATCATTAACCACTTCCATCCCTTACTCTCA 957
 DB 1028 AACCGCCAGAGTGTAGTGTCCCTCATGTAACCACTTATACCACTTTATCCCTTATGAACA 1087
 QY 958 AATGCTGAATTTGGAAGACGATCGCTGTTATTAATCCCTTTCGTTATCGTTCAACCA 1017
 DB 1088 AATGCTGAATTTGGAAGAACGAATTCGTATTAATTTCCCTTTCGTTATCGTTCAACCA 1147
 QY 1018 TTGGGTACCAGATTCAGGCCAGAACCAACAGTCCACAAACGACTCCCGGAACCTAGTCC 1077
 DB 1148 TTGGGTACCAGATTCAGAACCCAGAACCAAGTCCAGATTCGACTCCGGAACTTAGTCC 1207
 QY 1078 AGCCCGCAACCTGCACCAAAATCTTTAAATAGACTCAAAATTTTC-----TTT 1125
 DB 1208 AAGTCTGCAACCTGCACCAAAATCTTCAACAGCTCCCAAGCAATCCAATTTGATGAGAAAT 1267
 QY 1126 GCTTACTAGCTGTTACGAAAGTTGGGGAAGATATGTTTGAAGAAAGGGCATCTC 1185
 DB 1268 GGTCAAAAGAGTGTTCGAAAGATAGGCGATGTTATGTTTGGAGGAAATGGAGTTTC 1327
 QY 1186 TCGTTATGCTTTTGGGAAAGATTTTACCACTCTCAAACTGTTAAATCTTTGAAAGCAAGTT 1245
 DB 1328 TCGTTATATCCAGCCAGGATCTTTTTCAGCAGAAACAGCAGCAGGATTTGATGAAACT 1387
 QY 1246 ATCAAAACAGAGAGTGTTCACACATTTAACTGCTAAAAAGAAAAATTTGTTCTCTCG 1305
 DB 1388 GGCAAGCAGGAAAGTTTATCTCTAAAGCTAGGAGCTAAGAAACCTGACCTCCCATCTAG 1447
 QY 1306 TCACCAAGAAATTTATGATAAGCATATACTGTTAACTGAGGCTCAATAAGCCCTGTT 1365
 DB 1448 TGATCGAGAAATTTTCAATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1507
 QY 1366 TGNAAATAGGCTCGTAATTTCTGATTTTCAAGCCTTTAGACAAATTTATAGAACGTTGAA 1425
 DB 1508 TGATTAATAAGTTCAGCAAGTTGATTTTGAAGTTTGGATTAACCTGTTGGAACGACTCA 1567
 QY 1426 TGATGAATCGATTAATAAGAAAAATTTGATGATGATTTTGGCATTTCTAGACCAAT 1485
 DB 1568 GGATGCTCAAGTGAATAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGCTCCGAT 1627
 QY 1486 TACCCATCCAGAGCAGTTCGCAACCAAAATTTCTCAATTTGAGTATACCTGAGACGAACT 1545
 DB 1628 TCGTATCCAGAACGTTTAGGAAACCAAAATGCGCAAAATTAACCTACACTGATGATGAT 1687
 QY 1546 TCGTATTTGCTCAATTAGCTGATAAGTATATAACAGCTGAGATGTTTATCATTTTGTGAAACA 1605
 DB 1688 TCAAGTAGCAAGTTGCGCAGGCAAGTACACACAGAGACGCTTATATCTTTGATCCTCG 1747
 QY 1606 TGATATAATCAGTATGAAGAGATGATATGTAACGCTCATATGGGCAATAGTCACTG 1665
 DB 1748 TGATATAACAGTATGAGGGGATGCTTATGTAATCTCCATATGACCATAGCCACCTG 1807
 QY 1666 GATTGGAAAGATAGCTTTCTGATAAGGAAAGTTCAGCTCAAGCTCAAGCTATACCTAAGA 1725
 DB 1808 GATTAAAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCCAGGCTTATGCTAAGA 1867
 QY 1726 AAAAGTATCTTACCTTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAATTTGGAGATAG 1785
 DB 1868 GAAAGTTTGACCTCTCTTCCAGAGACCAACAGGATTCAGGAAATCTGAGGCAAAAG 1927
 QY 1786 TGCAGCAGCTATTTTCAATTCGTTGTAAGGGGAAAAAGCAATTCACCTCGTTTCGACTTCC 1845
 DB 1928 AGCAGAGCTATCTTACAAACCGCTGAAAGCAGCTAAGAGGTGCCACTTGTATCGTATGCC 1987

1846 ATATGTTGAGCATACAGTTCAGGTTAAACCGGTAATTTGATTATTCCTCATAGGA 1905
 1888 TTACAATCTTCAATATACTGTAGAGTCAAAACGGTAGTTTAAATCATACCTCATATGA 2047
 1906 TCATTACCAATAATTAATTTGCTTGGTTTGATGATCACAATACAAAGCTCCAAATGG 1965
 2048 CCATTACCAATAACATCAAAATTTGAGTGGTTTGACGAAGGCGCTTTATGAGGCACCTAAGG 2107
 1966 CTATACCTTGAAGATTTGTTTGGACGATTAAGTACTAGTAGACACCTGACGAACG 2025
 2108 GTATAGTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAC 2167
 2026 TCACATTTCTAATGATCGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCA 2085
 2168 TCCGCAATTCAGTAATGTTTGTGTAACGCTAGTGACCATGTTTCGTAAATAATAGGCAGA 2227
 2086 CAGTGAAGATCCCAATTAAGAACTTCAAGCGGATGAAGAGCCAGTACGAGGAACACCTGC 2145
 2228 CCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAAGTGAAGCAACTCACCC 2287
 2146 TGAGCCAGAAG 2156
 2288 TGAATCTGATG 2298

RESULT 24
 AAX25394
 ID AAX25394 standard; DNA; 2163 BP.

XX AC AAX25394;

XX DT 19-JUL-1999 (first entry)

XX DE Streptococcus pneumoniae complement C3-degrading protease DNA.

XX KM Human complement C3-degrading protease; vaccine; infection; meningitis;
 XX KW pneumonia; xerotransplantation; transplant rejection; inflammation; ds.

XX OS Streptococcus pneumoniae.

XX PN W0915675-A1.

XX PD 01-APR-1999.

XX PF 24-SEP-1998; 98WO-US020186.

XX PR 24-SEP-1997; 97US-0059907P.

XX PA (MINU) UNIV MINNESOTA.

XX PA (AMCY) AMERICAN CYANAMID CO.

XX PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;

XX DR WPI; 1999-254719/21.

XX DR P-PSDB; AAY05753.

XX PT New isolated human complement C3-degrading proteinase.

XX PS Claim 54; Page 52-54; 56pp; English.

XX CC This DNA sequence encodes a 79 kDa protein (see AAY05753) of
 CC Streptococcus pneumoniae serotype 4 that is capable of degrading human
 CC complement protein C3 (HCP3). It was identified in the *S. pneumoniae*
 CC serotype 4 genome by alignment to another novel open reading frame (see
 CC AAX25393) that codes for a 20 kDa HCP3 protease (AAY05752) of *S.*
 CC pneumoniae serotype 3. This suggested the open reading frame that codes
 CC for the 20 kDa protein may be part of a larger open reading frame. Amino
 CC acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence
 CC identity with amino acids 170-227 and 258-300 of the 79 kDa protein.
 CC proteins and peptides or polypeptides containing these regions, and DNA
 CC sequences encoding them (nucleotides 507-681 and 827-999 of the present
 CC sequence) are claimed. HCP3 proteases and polypeptides can be used as
 CC immune system stimulating compositions (claimed). They can produce an

CC immune response against *S. pneumoniae* to immunize or treat a mammalian
 CC subject against infection or colonization (claimed). They can produce a B
 CC cell response, a T cell response, an epithelial cell response, or an
 CC endothelial cell response (claimed). The expression of the proteins on
 CC the surface of an organ of an animal used in xenotransplantation can be
 CC used to inhibit C3-mediated inflammation and rejection
 XX SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 U; 0 Other;

Query Match 36.4%; Score 869.4; DB 2; Length 2163;
 Best Local Similarity 65.2%; Pred. No. 3.5e-209;
 Matches 1436; Conservative 0; Mismatches 627; Indels 141; Gaps 5;

QY 254 ATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTAATGAGGTCAAGGGTGA 313
 Db 1 ATGAAGATCCGAATATATCATGTTCAAGGATTCAGACATGTCATGAATCAAGGGTGT 60
 QY 314 TATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGTCGCCAGCGGAT 373
 Db 61 TATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGTCAGCTCATGCCGAT 120
 QY 374 AACGTCGCTACAAAGAGGAAATCAATCGACAAAACAAGAGCATAGTCAACATCGTGA 433
 Db 121 AATATTCGACAAAAGAGAGATTAAACGTCAGAGCGGAACACAGTCTATATACGGG 180
 QY 434 GGTGGAATCTCCAAAGAAACGATGGTGTGTTGCTTGGCAGCTTTCGCAAGGACGCTATACT 493
 Db 181 GGTGGTTCT-----AACGATCAAGCAGTAGTTGCGCCAGAGGCCAAGCAGCTATACA 234
 QY 494 ACAGATGATGGTTATATATCTTTAATGTTCTGATATCATAGAGGATCTGGTGATGCTTAT 553
 Db 235 ACGGATGATGGTTATATCTTTCAATGATCTGATATCATTTGAGGACACGGGTGATGCTTAT 294
 QY 554 ATGTTCTCTCATGAGATCATTAACATTACATTCCTAAGAAATGATTTATCAGCTAGCGAG 613
 Db 295 ATGTTCTCTCAGCGGACCATTAACATTACATTCCTAAGAAATGATTTATCAGCTAGCGAG 354
 QY 614 TTGGCTCTCAGAAAGCCTTCTCTATCTGTCGAGGAAATCTGTCAAATTCAAAGAACCTAT 673
 Db 355 TTAGTCTCTCAGAAAGCCTTAT-----GGAAAT 381
 QY 674 GCGGCAAAAATAGCGATATAACATTTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCA 733
 Db 382 GGGAAAGCAGGGATCTCGTCTCTTCAAGTTCTAGTTTATATGCAAAATCCAGCTCAACCA 441
 QY 734 GGAATACAAATCTAAACAAGCAACAACAGCAACTAACAGTCAAGCAAGTCAAAAGT 793
 Db 442 AGATTGTCAGAGAACCAATCTGACTGTCTACTCAACTTA---TCATCAAAATCAAGGG 498
 QY 794 AATGACATTGATAGTCTCTTGAACACAGCTCTACAAACTGCTTTGAGTCAACGACATGTA 853
 Db 499 GAAACACATTTCAAGCCTTTTACGTTGAATTTGATGCTAAACCTTATCAGAACGCGCATGTG 558
 QY 854 GAATCTGATGGCTTGTCTTTGATCGACCAATACAGTCAAGTCAAGTCAAGTCAAGTGT 913
 Db 559 GAATCTGATGGCTTATTTTCGACCCGCAATCAAGTCAAGTCAAGTCAAGTCAAGTGT 618
 QY 914 GCAGTGCACACGGAGATCATTAACACTTCTCCCTTACTCTCAAATGTCTGAATTCGAA 973
 Db 619 GCTGTCCCTCATGTTAACCATTACCACTTTATCCCTTATGNAACAAATGCTGAATTCGAA 678
 QY 974 GAACGAATCGCTCGTATTTATTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 1033
 Db 679 AAACGAATTCGCTCGTATTTATTTCCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 738
 QY 1034 AGGCGAAGCAACAGTCCACACCGCACTCCGGAACCTAGTCCAGCGCCGCAACCTGCA 1093
 Db 739 AGACCAAGCAACCAAGTCCCAATCGCACTCCGGAACCTAGTCCAGTCCGCAACCTGCA 798
 QY 1094 CCAATCTTAAATAGACTCAA-----ATTCTTTCTTTGGTTAGTCACTGGTA 1141
 Db 799 CCAATCTCAACCGCTCCAAGCAATCCAATTCATGAGAAATTTGGTCAAGAAGAGCTGTT 858

1142 CGAAGAGTCTGGGAGGATATGTTATCGAAGAAAAGGCGATCTCTCGTTATGCTTTGGC 1201
 859 CGAAGAGTGGCGATGGTTATGCTTTGAGGAGATGGAGTTCTCGTTATATCCAGCC 918
 1202 AAGAGTTTACATCTGAACTGTAAATCTTTGAAAGCAAGTTATCAAAACCAAGAGAGT 1261
 919 AAGGATCTTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 978
 1262 GTTTCACACACTTTTACTGCTAAAGAAAAGAAATGTTGCTCTCGTGACCAAGAAATTTAT 1321
 979 TTATCTCATAGCTAGGAGCTAAGAACTGACCTCCCATCTAGTGTATGATCGAGAAATTTAC 1038
 1322 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGAGCTTTGTTGNAATAAGGTCGT 1381
 1039 AATAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACTTGATATAAGGTGGA 1098
 1382 AATCTGATTTCAAGCCTTTAGCAAAATTTATGAAAGCTTTGAATGATGAATCGACTAAT 1441
 1099 CAAGTTGATTTGAGGCTTTGGATAACCTGTTGGAAGGAGCTCAAGGATGTCCCAAGTAT 1158
 1442 AAGAAAATTTGATAGATGATTTATTTGGCATTTCTAGCACCATTACCCATCCAGAGCGA 1501
 1159 AAGTCAAGTTAGTGTATGATATCTTGGCTTTCTTAGCTCCGATTCGTATCCAGAACGT 1218
 1502 CTTGGCAAAACCAAAATCTCAAAATTTGATATCTGAAAGCAAGCTTCGTATTTGCTCAATTA 1561
 1219 TTAGGAAAACCAAAATCTGCAAAATTTACTACATGATGAGATTCAGTAGCCAAGTTG 1278
 1562 GCTGATAAGTATACAAAGCTGAGATGTTTATTTTGAATGAAATCATGATATTAATCAAGTAT 1621
 1279 GCAGGCAAGTATACAAAGCAAGAGCGGTTATCTTTGATCTCTGATATTAACCAAGTAT 1338
 1622 GAAGGAGATGATATGATGAGGCTCATATGAGGCAATGCTGAGTATGGAAGAGATAGC 1681
 1339 GAGGGGATGCTTATGATCTCACTTCCATATGACCATAGCCATGGAATTAAGAAAGATAGT 1398
 1682 CTTTCTGATAAGGAAAAGTTCAGCTCAAGCTATCTAATAAGAAAAGTATCTTACCT 1741
 1399 TTGCTGAGCTGAGAGCGGAGCGGAGGCTTATGCTTAAGAGAAAAGTTTGACCCCT 1458
 1742 CATCTCCAGAGCAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTAC 1801
 1459 CTTTCAGACAGCATCAGGATTCAGGAAATCTAGGCAAAAGGAGCAGAGCTATCTAC 1518
 1802 AATCTGTGAAAGGGAAGAAACGAATTCCTACCTGCTGACCTCCATATATGTTGAGCAT 1861
 1519 AACCGGTGAAGAGCTAAGAGGTCGACCTGATGCTGATGCTTACAACTCTTCAATAT 1578
 1862 ACAGTTGAGGTTAAAAACCGTAAATTTGATTTATCTCTCAATAAGGATCATTTACCATATAT 1921
 1579 ACTGTAGAGTCAAAAACCGTAGTTTAACTACCTCATTTATGACCATTTACCATATATC 1638
 1922 AATTTGCTGTTGATGATCACATACATCAAGCTCCAAATGGCTTATACCTTTGGAAGAT 1981
 1639 AATTTGAGTGTGTTGACGAAGGCTTTTATGAGGACCTTAAGGGGTATATCTCTGAGGAT 1698
 1982 TTGTTTGGCAGCATTAAGTACTAGTACGATAGACACCTGAGGAGGCTCCAAATTTATGAT 2041
 1699 CTTTGGGCACTGTCAAGTACTATGTGCAAACTCCAAAGCAAGCTCCGATTTGAGATAT 1758
 2042 GGATGGGCAATGCCAGTGTGATGATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAAT 2101
 1759 GGTGTTGTTGATGCTAGCGACCATGTTTCAAGAAACAAAATGTTCAAGCTGATACCAT 1818
 2102 AAGAATTTCAAGCGGATGA----- 2121
 1819 CAAACGGAAGAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAACCCCT 1878
 2122 -----AGAGCCAGTAGAG 2134
 1879 CGAAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAAACCAACAGAGGAAACGAGAA 1938
 2135 GAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTC 2194

1939 GAATCCAGAGGAATCAGAAGAACTCAGGTCGAGACTGAAAAGGTTGAAGAAAACCTG 1998
 2195 AAAGAAGCAGAGGATTTTGTCTTGGAAAAGTAAACGGATTCTAGTCTGAAAAGCAATGCAACA 2254
 1999 AGAGAGGCTGAAGATTTACTTGGAAAATCCAGGATCCAAATTTCAAGTCCATGCCAAA 2058
 2255 GAAACTCTAGCTGGTTTACGAAATTAATTTGACTCTTCAAAATTTGATGATCAATATGATC 2314
 2059 GAGACTCTCAGAGGATTAATAATTTACTATTTGGCACCAGGACCAACATACTATT 2118
 2315 ATGGCAGAGCAGAAAATTTACTTGGCTTCTTAAAGGAGTAA 2358
 2119 ATGGCAGAGCTGAAAACCTATTGGCTTTTATTAAAGGAGATTA 2162

RESULT 25

AAV52376

ID AAV52376 standard; DNA; 2359 BP.

XX AAV52376;

XX 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:243.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

XX computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019588.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;

XX Dougherty BA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

XX polynucleotide sequences - useful in diagnostic kits and assays, and

XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 1265-1266; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridise to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 11:19:10 ; Search time 6082 Seconds
(without alignments)
11729.817 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTGGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	3.0	906	28	BH153606
2	69.8	2.9	890	28	BH146886
3	67.8	2.8	180	28	AF325819
4	67.8	2.8	908	28	AZ548467

					BH153606 ENTSS9TTP
					BH146886 ENTPK48TF
					AF325819 AF325819
					AZ548467 ENTEK30TR

C 5	66.4	2.8	829	29	BX173672
C 6	64.8	2.7	843	28	AZ551618
C 7	63.8	2.7	877	28	AZ531291
C 8	62.2	2.6	912	28	AZ551092
C 9	60.8	2.5	849	28	AZ546009
C 10	59.2	2.5	931	28	BH160272
C 11	58.6	2.5	1101	29	CNS0039G
C 12	57.2	2.4	900	28	AZ549980
C 13	56.8	2.4	905	28	AZ550256
C 14	56.6	2.4	645	12	BM165350
C 15	56.6	2.4	747	12	BM162732
C 16	55.2	2.3	816	28	AZ535744
C 17	55	2.3	671	12	BM163120
C 18	55	2.3	647	12	BM160252
C 19	55	2.3	694	12	BM168242
C 20	55	2.3	717	12	BM160500
C 21	55	2.3	756	12	BM162492
C 22	55	2.3	794	12	BM159906
C 23	55	2.3	795	28	AZ528485
C 24	55	2.3	823	28	AZ576218
C 25	55	2.3	843	28	BH139532
C 26	53.8	2.2	721	13	BM141179
C 27	53.6	2.2	656	13	BM144040
C 28	53.6	2.2	844	29	BX139987
C 29	53.2	2.2	752	14	CD099026
C 30	52.8	2.2	671	13	BW273704
C 31	52.8	2.2	890	28	AZ530768
C 32	52.4	2.2	677	12	BM161314
C 33	52.4	2.2	725	12	BM169486
C 34	52.4	2.2	866	28	AZ527885
C 35	52.4	2.2	976	28	BH149983
C 36	52.2	2.2	945	29	CNS008CA
C 37	51.8	2.2	435	12	BM169710
C 38	51.8	2.2	880	28	AZ529191
C 39	51.6	2.2	853	29	CNS023KH
C 40	51	2.1	500	9	AU088119
C 41	50.8	2.1	726	28	AZ196050
C 42	50.2	2.1	1007	29	CNS06X9S
C 43	50	2.1	445	12	BM167469
C 44	50	2.1	555	13	BH496963
C 45	49.8	2.1	707	12	BM164483
C 46	49.6	2.1	598	12	BM170666
C 47	49.6	2.1	605	12	BM163520
C 48	49.4	2.1	629	12	BJ073162
C 49	49.4	2.1	687	12	BM169665
C 50	49	2.1	1044	13	BX415231
C 51	48.8	2.0	1101	29	CNS0022U
C 52	48.6	2.0	605	12	BM162501
C 53	48.6	2.0	861	13	BW266138
C 54	48.4	2.0	546	13	BH497202
C 55	48.4	2.0	891	28	AZ683582
C 56	48.4	2.0	960	28	CC143364
C 57	48.2	2.0	468	12	BG553005
C 58	48.2	2.0	537	13	BH498263
C 59	48.2	2.0	554	12	BM275335
C 60	48.2	2.0	648	10	BH349803
C 61	48.2	2.0	661	9	AU060703
C 62	48.2	2.0	679	12	BH440521
C 63	48.2	2.0	756	12	BM165937
C 64	48.2	2.0	782	13	BM117562
C 65	48	2.0	548	12	BM159818
C 66	48	2.0	1101	29	CNS00LT2
C 67	47.8	2.0	713	13	BH412428
C 68	47.8	2.0	718	13	BH257897
C 69	47.8	2.0	773	13	BH249618
C 70	47.8	2.0	792	13	BH114945
C 71	47.8	2.0	799	13	BH332339
C 72	47.8	2.0	834	12	BH090507
C 73	47.8	2.0	884	13	BH137322
C 74	47.6	2.0	880	28	AZ669474
C 75	47.4	2.0	541	29	CNS04KUS
C 76	47.4	2.0	731	13	BM115605
C 77	47.2	2.0	311	14	CA699402

Db	615	ACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAG	674
Qy	2157	TCCTCAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAAGAAGCAGCAAGTT	2209
Db	675	ACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGATGAAGAAGACGAT	727
RESULT 3			
LOCUS	AF325819	180 bp	DNA linear GSS 08-MAR-2002
DEFINITION	AF325819	Leptospira interrogans serovar lai DNA	Leptospira interrogans serovar lai genomic clone 18, genomic survey sequence.
ACCESSION	AF325819		
VERSION	AF325819.1	GI:19263459	
KEYWORDS	GSS.		
SOURCE	Leptospira interrogans serovar lai		
ORGANISM	Leptospira interrogans serovar lai		
REFERENCE	Bacteria; Spirochaetes; Spirochaetales; Leptosiraceae; Leptospira.		
AUTHORS	1 (bases 1 to 180)		
TITLE	Hu, C. and Bao, L.		
JOURNAL	Leptospira interrogans serovar lai DNA		
COMMENT	Unpublished (2002)		
	Contact: Hu C		
	Leptospirosis Research Unit		
	West China University of Medical Sciences		
	17# Renmin Nan Road, Chengdu, Sichuan 610041, P.R. China		
	Email: huchanghua@263.net		
	Class: unknown.		

```

FEATURES
source
      1. .180
      /location=Qualifiers
      /organism="Leptospira interrogans serovar lai"
      /mol_type="genomic DNA"
      /db_xref="taxon:57678"
      /clone="18"
      /clone_lib="Leptospira interrogans serovar lai DNA"
      /notes="obtained through DNA subtraction using the
nonpathogenic strain Leptospira biflexa serovar Patoc
strain Patoc 1"

```

	Query Match	2.8%	Score 67.8	DB 28	Length 180
	Best Local Similarity	62.9%	Pred. No. 1.8e-05		
	Matches 105	Conservative 0	Mismatches 52	Indels 0	Gaps 0
Qy	249	TACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAGGATATTGTTTAATGAGTCAAGG	308		
Db	1	TATTAATGACGGACCCCAATTATCAGTTTAAACAGTCTGATGTTGTTTAAAGAGATTCCTAG	60		
Qy	309	GTGGATATGTTTCAAGGTAGATGGAAAATACTATGTTTACCTTAGGATGTCGCCCAAG	368		
Db	61	ATGGCTATATTATCAAGGTTGATGGCAAGTATTATGTTTATCTAAAGCCAGGGAAGCAAGC	120		

AZ548467	908 bp	DNA	linear	GSS 14-NOV-2000
ENTR3K30TR	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
genomic,	genomic survey sequence.			
AZ548467				
ACCESSION				
VERSION	AZ548467.1	GI:11172102		
KEYWORDS	GSS.			
SOURCE	Entamoeba histolytica			
ORGANISM	Entamoeba histolytica			
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.			
AUTHORS	1 (bases 1 to 908)			
TITLE	Loftus, B., Van Aken, S. and Fraser, C.			
JOURNAL	Determination of clone end sequences from Entamoeba histolytica			
	HM1:IMGS sheared DNA library			
	Unpublished (2000)			

TITLE Determination of clone end sequences from *Entamoeba histolytica*

JOURNAL HMI:IMSS sheared DNA library
Unpublished (2000)


```

High quality sequence start: 22
High quality sequence stop: 829.
FEATURES
    source
        1. .877
            /organism="Entamoeba histolytica"
            /mol_type="genomic DNA"
            /strain="HM1:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="vector: pHO1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Borell, Oxford University Press, 1999)."
ORIGIN
    Query Match      2.7%; Score 63.8; DB 28; Length 877;
    Best Local Similarity 43.4%; Pred. No. 0.00029;
    Matches 293; Conservative 0; Mismatches 382; Indels 0; Gaps 0;
QY 1535 GAAGCAGAGTTCGATTGCTCAATTAGCTGATGAAGTATACACGTCAGATGGTTACATT 1594
DB 785 GAAGCAGATGATGATGATGATGAAGAGAGATGATGAAGATGATGATGATGAAGAA 726
QY 1595 TTTGATGAACATGATATTAATCAGTGTGAGGAGATGATGAAGCGCTCATATGGC 1654
DB 725 GATGATGATGAAGAGATGATGATGAAGAGATGATGATGAAGAGATGATGAAGAA 666
QY 1655 CATAGTCACTGGATTGGAAGAGATAGCCTTTCTGATGAAGGAAAAGTTGACGCTCAAGCC 1714
DB 665 GATGATGATGAAGAGAGATGATGATGAAGATGATGAAGAGATGATGAAGAGAGAA 606
QY 1715 TATCTAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1774
DB 605 TATGAAATTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 546
QY 1775 ACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1834
DB 545 GATGAAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 486
QY 1835 GTTCGACTTCCTATATATGTTGAGCATACAGTTGAGGTTAAAGCGGTAATTTGATTATT 1894
DB 485 GAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAA 426
QY 1895 CCTAATAGGATCAATACATTAATTAATTTGCTGTTGATGATGATGATGATGATGATGATGAT 1954
DB 425 GAAGATGATGAAGATGATGATGAAGAGAGATGATGATGAAGATGATGAAGATGATGAAGACAT 366
QY 1955 GCTCAATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2014
DB 365 GAAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 306
QY 2015 CCTGACGACGCTCCACATTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2074
DB 305 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAA 246
QY 2075 AAGAAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2134
DB 245 GATGAAGATGATGATGATGAAGAGAGATGATGATGAAGAGAGATGATGAAGAGAGAGAT 186
QY 2135 GAACACCTGCTGAGCGACGAGTCCCTCAAGTAGAGATGATGAAGAGAGATGATGAAGAGAGAT 2194
DB 185 GATGAAGAGATGATGAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAA 126
QY 2195 AAGAAGCAGAGATT 2209

```

```

Db 125 GATGATGAAGAGAT 111
RESULT 8
AZ551092
LOCUS
DEFINITION
    AZ551092 912 bp DNA linear GSS 14-NOV-2000
    ENTJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
    genomic, genomic survey sequence.
ACCESSION
    AZ551092 GI:11176393
VERSION
    AZ551092.1
KEYWORDS
    GSS.
SOURCE
    Entamoeba histolytica
    Eukaryota; Entamoebidae; Entamoeba.
ORGANISM
    1 (bases 1 to 912)
    Loftus, B., Van Aken, S. and Fraser, C.
    Determination of clone end sequences from Entamoeba histolytica
    HM1:IMSS sheared DNA library
    Unpublished (2000)
    Contact: Brendan J Loftus
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0208
    Fax: 301 838 3543
    Email: bjl@loftus.tigr.org
    Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
    DNA library
    Seq primer: M13-Forward
    Class: Shotgun
    High quality sequence start: 17
    High quality sequence stop: 861.
FEATURES
    source
        1. .912
            /organism="Entamoeba histolytica"
            /mol_type="genomic DNA"
            /strain="HM1:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="vector: pHO1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Borell, Oxford University Press, 1999)."
ORIGIN
    Query Match      2.6%; Score 62.2; DB 28; Length 912;
    Best Local Similarity 43.3%; Pred. No. 0.00071;
    Matches 292; Conservative 0; Mismatches 383; Indels 0; Gaps 0;
QY 1535 GAAGCAGAGTTCGATTGCTCAATTAGCTGATGAAGTATACACGTCAGATGGTTACATT 1594
DB 213 GAAGAGATGATGATGATGATGAAGAGAGATGATGATGAAGATGATGATGATGATGATGAAGAA 272
QY 1595 TTTGATGAACATGATATAATCAGTGTGAGGAGATGATGATGAAGATGATGATGAAGATGATGAAGAA 332
DB 273 GATGATGATGAAGAGATGATGATGAAGAGAGATGATGATGAAGATGATGATGAAGAA 1654
QY 1655 CATAGTCTCGATTGGAAGAGATAGCCCTTTCTGATGAAGAGAGATGATGATGATGATGATGATGAT 1714
DB 333 GATGATGAAGAGAGAGATGATGATGAAGAGATGATGAAGATGATGAAGAGATGATGAAGAGAGAA 392
QY 1715 TATCTAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1774

```

Db 393 TATGAATTAGAAGATGATGATGATGAAGAAGAAGATGATGATGAAGATGATGAAGAC 452
 QY 1775 ACTGGAGATAGTCACAGCTATTTCACATCGTGTGAAGGGGAAAAACGAATTCACATC 1834
 Db 453 GATGAAGACACCAATATGAATTTAGAGATGATGATGATGAAGAAGAAGATGATGAT 512
 QY 1835 GTTCGACTTCCATATATGTTGAGCATACAGTTGAGGTTTAAACCGTAAATTTGATTATT 1894
 Db 513 GAAGAAGAAGAGATGATGATGAAGATGATGATGAAGAAGAAGATGATGAAGATGAAGAA 572
 QY 1895 CTTCAATAGCATCATTAACATATATTAATTTGTTGGTTTGGATGATCACACATACAA 1954
 Db 573 GAAGATGATGAAGATGATGATGAAGAAGAAGATGATGATGAAGATGATGAAGACGAT 632
 QY 1955 GCTCCAAATGGCTATACCTTGGAGATTTCTTGGACGATTAAGTACTACGTAGAACAC 2014
 Db 633 GAAGACGACGAATATGAATTTAGAGATGATGATGAAGAAGAAGATGATGATGAAGAA 692
 QY 2015 CTTGACGAACGCTCCACATTTCTAATGATGGATGGGCAATGCCAGTCAGCATGTTTAGGC 2074
 Db 693 GAAGAAGATGATGATGAAGATGATGATGAAGAAGAAGATGATGATGAAGAGAA 752
 QY 2075 AGAAGAACCACAGTGAAGATCCAAATAGAACTTCAAAGCGGATGAAGAGCCAGTAGAG 2134
 Db 753 GATGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGAAGAAGAAGAT 812
 QY 2135 GAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGMAAAGTAGAAGCCCACTC 2194
 Db 813 GATGAAGAAGATGATGAAGAAGAAGATGATGATGAAGAAGATGATGAAGAAGATGATGA 872
 QY 2195 AAAGAAGCAGAAGATT 2209
 Db 873 GATGATGAAGAAGAT 887

RESULT 9
 AZ546009/c 849 bp DNA linear GSS 14-NOV-2000
 LOCUS ENTFW53TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.

ACCESSION AZ546009
 VERSION AZ546009.1 GI:11167130
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 849)
 Loftus,B., Van Aken,S. and Fraser,C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)

JOURNAL
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Claes: shotgun
 High quality sequence start: 26
 High quality sequence stop: 796.

FEATURES
 source
 1..849
 Location/Qualifiers
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, I.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Query Match 2.5%; Score 60.8; DB 28; Length 849;
 Best Local Similarity 43.4%; Pred. No. 0.0015;
 Matches 275; Conservative 0; Mismatches 358; Indels 0; Gaps 0;
 QY 1190 TATGCTTTTGGCAAGATTTACCATCTGAAACTGTTTAAATAATCTTGAAGCAAGTTATCA 1249
 Db 742 TTGACTTAGTGAATAATGTAGAAATGGAACATGATACCGTTTGAAGACATGATGAATTA 683
 QY 1250 AAACAAGAGAGTGTTCACACACTTTTAACTGCTTAAAAAAGAAAAATGCTCTCCTCGTAC 1309
 Db 682 GAAGAAGATGATGATGAATTTGAAATTAGAAGACGAAGATGATGATGATGATGATGAT 623
 QY 1310 CAAGAATTTTATGATAAGCATATATCTGTTAACTGAGGCTCTAAAGCCCTTTGTTGNA 1369
 Db 622 GACGAAGAAGATGATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
 QY 1370 AATAAGGGTCTGTAATTTCTGATTTCCAGGCTTTAGACAAATTTATAGAACGCTTGTATGAT 1429
 Db 562 GAATTTGAATTAGAAGATGAAGAAGATGAAGATGATGATGATGATGATGATGATGATGATGAT 503
 QY 1430 GAATCGACTTAATAAGAAAAATTTGGTAGATGATTTATTGGGCATTTCTAGCACCATTACC 1489
 Db 502 GAAGATGATGATGAAGATGAAGAAGATGATGATGATGATGAAGAAGACGAAGAAGATGAT 443
 QY 1490 CATTCAAGAGCGACTTTGGCAACCAATTTCTCAATTTAGTATCTAGAGACGAAAGTTGCT 1549
 Db 442 GATGATGATTTGAATTTAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGATGATGAT 383
 QY 1550 ATTGCTCAATTAGCTGATAAGTATACACGTCAGATGTTTACATTTTGTGATCAACATGAT 1609
 Db 382 GATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGAT 323
 QY 1610 ATAATCAGTGAAGAGATGCATATATGAAGCCCTCATATGGGCCATAGTCACCTGGATT 1669
 Db 322 GAAGATGATGATGATGATGATGAAGATGAAGAAGAAGATGATGATGAATTTGAAATTA 263
 QY 1670 GGAAGAAGATGACCTTTCTGATAAGAAAAAGTTGAGCTCAAGCCTTACTTAAGAAAAA 1729
 Db 262 GAAGATGAAGAAGATGAAGATGATGATGAAGAAGACGAAGAAGATGATGATGAATTT 203
 QY 1730 GGTATCTTACCTCCATCTCCAGACGACATGTTAAAGCAAAATCCAACTGGAGATAGTCA 1789
 Db 202 GAATTAGAAGATGAAGATGAAGATGAAGATGAAGAAGATGAAGATGATGATGAAGAAGAA 143
 QY 1790 GCAGCTATTTACAATCGTGTGAAGGGGAAAAA 1822
 Db 142 GAAGAAGAAGATGATGATGAAGATGAAGAAGAA 110

RESULT 10
 BHI60272
 LOCUS ENTQV49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION BHI60272
 VERSION BHI60272.1 GI:15733710
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoebidae; Entamoeba.

Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 100
 High quality sequence stop: 872.
 Location/Qualifiers
 1. 905
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IM5"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Query Match 2.4%; Score 56.8; DB 28; Length 905;
 Best Local Similarity 44.8%; Pred. No. 0.014;
 Matches 269; Conservative 0; Mismatches 323; Indels 9; Gaps 1;
 1139 GTACGAAAGTTGGGGAAGGATATGATTCGAAAGAAAGGCGATCTCTCGTTATGCTTT 1198
 Db 780 GAATTAGAGAGATGATGATGATGAATTTGAATTAGAGAGAGATGATGATGAT 721
 Qy 1199 GCGAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAG 1258
 Db 720 GATGAGACGGAAGAGATGAAGAGAGATGATGATGATGATGATGATGATGATGAT 661
 Qy 1259 AGTGTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGACCAAGATTT 1318
 Db 660 GATGATGAATTTGAATTAGAGATGAAGAGATGAAGATGATGATGATGATGATGAT 601
 Qy 1319 TATGATTAAGCATATAATCTGTTAACTGAGGCTCATATATTTGGCATTCCTAGCCCAATACC 1369
 Db 600 GATGATGAAGATGATGATGAAGATGAAGAGATGATGATGATGATGATGATGATGAT 541
 Qy 1370 AATAAGGTCCTAATTCGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGAT 1429
 Db 540 GATGATGATGATGAATTTGAATTAGAGATGAAGAGATGAAGAGATGAAGAGATGAT 481
 Qy 1430 GAATCGACTAATAAGAAATTTGGTAGATGATGATTTATTTGGCATTCCTAGCCCAATACC 1489
 Db 480 GATGATGAAGATGATGATGAAGATGAAGAGATGATGATGATGATGATGATGATGAT 421
 Qy 1490 CATTCAGAGCGACTTGGCAACCAAAATTCCTCAAAATTTAGTATGATGATGATGATGAT 1549
 Db 420 GATGATGATGATGATGATGAAGATGAAGAGAGATGATGATGATGATGATGATGAT 361
 Qy 1550 ATTGCTCAATTAGCTGATAGTATACACGTCAGATGATGATGATGATGATGATGATGAT 1609
 Db 360 GATGAAGAAGATGAAGATGATGATGAAGAGAGAGATGATGATGATGATGATGATGAT 301
 Qy 1610 ATAATCAGTGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
 Db 300 TTGAAGATGAAGATGAAGATGAAGATGAAGAGAGATGATGATGATGATGATGATGAT 241
 Qy 1670 GGAAGAATAGCCTTTCTGTAAGAAAGAGTTGAGCTTCAAGCCTTACTAAGAAAAA 1729
 Db 240 GAAGAAGATGATGATGAAGATGAAGAGAGATGATGATGATGATGATGATGATGATGAT 181
 Qy 1730 G 1730
 Db 180 G 180

RESULT 14

BM165350
 LOCUS
 DEFINITION
 EST567873 PyBS Plasmodium yoelii yoelii cDNA clone PYCML60 5' end,
 mRNA sequence.
 BM165350
 VERSION
 BM165350.1 GI:17311031
 EST.
 Plasmodium yoelii yoelii
 Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 645)
 Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Garucci,D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.
 Location/Qualifiers
 1. 645
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCML60"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /notes="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybrizAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pAD-GAL4) was
 excised from the HybrizAP vector and plasmid DNA
 isolated."

FEATURES

source

ORIGIN

Query Match 2.4%; Score 56.6; DB 12; Length 645;
 Best Local Similarity 47.4%; Pred. No. 0.014;
 Matches 170; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
 107 GATGAGGTTACCAAGCGTGAAGGAATCAATCTCTGAGCAAAATCGTCATCAAGATACAGAC 166
 Db 120 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179
 Qy 167 CAAGCTATGTCATTCACATGGCGACCATCATTTATTTACATGTTAAAGTTCTTAT 226
 Db 180 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239
 Qy 227 GACGCTATCATGATGAGGAATTACTCATGAAAAGATCCAAACTATTAAGCTAAAGATGAG 286

2.4%; Score 56.6; DB 12; Length 747;

[illegible]

Query Match 2.3%; Score 55; DB 12; Length 694;
 Best Local Similarity 47.1%; Pred. No. 0.035;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

107 GATGAGGTTAGCAAGCGTGAAGCAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 192 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 251
 |||||
 167 CAGGCTATGTCACCTCACATGCGGACACATATATTATCAATGCTGAAGTTCCTTAT 226
 |||||
 252 GACGATGATGACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
 |||||
 227 GAGCTATCATGCTGAGCAATTAATCATGAAACATCAAACTATAAGCTAAAAGATGAG 286
 |||||
 312 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
 |||||
 287 GATATTGTTAATGAGGTCAGGCGTGGATATGTTATCAAGGTPAGATGAGAAATCACTATGTT 346
 |||||
 372 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
 |||||
 347 TACCTTAAGGATGCTCCGACGCGATACCTGCTGATCAAAAGAGGAATCAATCGACAA 406
 |||||
 432 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
 |||||
 407 AAACAAGAGCATGCTCAACATCGTGAAGTGGAACTCAAGAAACGATGGTCTGTTGTC 465
 |||||
 492 GAGGATGAAGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 550
 |||||

RESULT 20
 BM160500
 LOCUS
 DEFINITION EST563023 PyBS Plasmodium yoelii yoe11i cdna clone PYCJS13 5' end,
 mRNA sequence.

ACCESSION BM160500
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Plasmodium yoelii yoe11i
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 717)
 Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1..717
 /organism="Plasmodium yoelii yoe11i"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoe11i"
 /db_xref="taxon:73239"
 /clone="PYCJS13"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 717;
 Best Local Similarity 47.1%; Pred. No. 0.036;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

107 GATGAGGTTAGCAAGCGTGAAGCAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 |||||
 205 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGATGAT 264
 |||||
 167 CAAGGCTATGTCACCTCACATGCGACACATATATTATCAATGCTGAAGTTCCTTAT 226
 |||||
 265 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
 |||||
 227 GACGCTATCATGCTGAGCAATTAATCATGAAACATCAAACTATAAGCTAAAAGATGAG 286
 |||||
 325 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
 |||||
 287 GATATTGTTAATGAGGTCAGGCGTGGATATGTTATCAAGGTPAGATGAGAAATCACTATGTT 346
 |||||
 385 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
 |||||
 347 TACCTTAAGGATGCTCCGACGCGATACCTGCTGATCAAAAGAGGAATCAATCGACAA 406
 |||||
 445 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
 |||||
 407 AAACAAGAGCATGCTCAACATCGTGAAGTGGAACTCAAGAAACGATGGTCTGTTGTC 465
 |||||
 505 GAGGATGAAGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
 |||||

RESULT 21

BM162492
 LOCUS
 DEFINITION EST565015 PyBS Plasmodium yoelii yoe11i cdna clone PYCKT28 5' end,
 mRNA sequence.

ACCESSION BM162492
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Plasmodium yoelii yoe11i
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 756)
 Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1..756
 /organism="Plasmodium yoelii yoe11i"
 /mol_type="mRNA"
 Location/Qualifiers

/strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCK28"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephadryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

Query Match 2.3%; Score 55; DB 12; Length 756;
 Best Local Similarity 47.1%; Pred. No. 0.036;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
 QY 107 GATGAGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 Db 113 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 172
 QY 167 CAAGCTATGTCACCTTCACATGCGACCACTATCATTTTACAAATGGTAAGGTTCTTAT 226
 Db 173 GACGATGATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGATGATG 232
 QY 227 GACGCTATCATCATGTCAGCAAGATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAG 286
 Db 233 GAAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
 QY 287 GATATTGTTAATGAGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAACTACTATGTT 346
 Db 293 GACGATGATGATGATGACGACGATGAAGATTTTGAAGATATGATGATGATGATGATGAT 352
 QY 347 TACCTTAAGGATGCTCCACCGGTAACGTCCTGACAAAAGAGGAATCAATCGACAA 406
 Db 353 GACGATGACGATGATGACGATGATGAAGACGATGATGAGGATGAAGATGATGAAGATGAT 412
 QY 407 AAACAAGAGCATAGTCAACATCGTGAAGTGGAACTCCCAAGAAACGATGCTGCTGTCG 465
 Db 413 GAGGATGAAGATGATGAAGAGATGACGATGATTTTAAAAAAGAGGTGGATCAATGC 471

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 756;
 Best Local Similarity 47.1%; Pred. No. 0.036;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
 QY 107 GATGAGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 Db 113 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 172
 QY 167 CAAGCTATGTCACCTTCACATGCGACCACTATCATTTTACAAATGGTAAGGTTCTTAT 226
 Db 173 GACGATGATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGATGATG 232
 QY 227 GACGCTATCATCATGTCAGCAAGATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAG 286
 Db 233 GAAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
 QY 287 GATATTGTTAATGAGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAACTACTATGTT 346
 Db 293 GACGATGATGATGATGACGACGATGAAGATTTTGAAGATATGATGATGATGATGATGAT 352
 QY 347 TACCTTAAGGATGCTCCACCGGTAACGTCCTGACAAAAGAGGAATCAATCGACAA 406
 Db 353 GACGATGACGATGATGACGATGATGAAGACGATGATGAGGATGAAGATGATGAAGATGAT 412
 QY 407 AAACAAGAGCATAGTCAACATCGTGAAGTGGAACTCCCAAGAAACGATGCTGCTGTCG 465
 Db 413 GAGGATGAAGATGATGAAGAGATGACGATGATTTTAAAAAAGAGGTGGATCAATGC 471

RESULT 22
 BM159906
 LOCUS
 DEFINITION EST562429 PyBS Plasmodium yoelii yoe11i cDNA clone PYCCK35 5' end, mRNA sequence.
 ACCESSION BM159906
 VERSION BM159906.1 GI:17305587
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoe11i
 ORGANISM Plasmodium yoelii yoe11i
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 794)
 Carleton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Garucci, D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished (2001)
 Contact: Jane Carleton
 Parasite Genomics Group

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carleton@tigr.org
 For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source
 1..794
 /organism="Plasmodium yoelii yoe11i"
 /mol_type="mRNA"
 /strain="17XL"
 /sub_species="yoelii"
 /db_xref="taxon:73239"
 /clone="PYCCK35"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephadryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 794;
 Best Local Similarity 47.1%; Pred. No. 0.037;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;
 QY 107 GATGAGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGAC 166
 Db 205 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGATGAAGACGATGAC 264
 QY 167 CAAGGCTATGTCACCTTCACATGCGACCACTATCATTTTACAAATGGTAAGTTCCTTAT 226
 Db 265 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 324
 QY 227 GACGCTATCATCAGTGAAGAAATTTACTCATGAAGATCCAAACTATAAGCTAAAGATGAG 286
 Db 325 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
 QY 287 GATATTGTTAATGAGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAACTACTATGTT 346
 Db 385 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
 QY 347 TACCTTAAGGATGCTCCACCGGTAACGTCCTGACAAAAGAGGAATCAATCGACAA 406
 Db 445 GACGATGACGATGATGACGATGATGAAGACGATGATGATGATGATGATGATGATGATGAT 504
 QY 407 AAACAAGAGCATAGTCAACATCGTGAAGTGGAACTCCCAAGAAACGATGCTGCTGTCG 465
 Db 505 GAGGATGAAGATGATGAAGAGATGACGATGATTTCTAAAAAAGAGGTGGATCAATGC 563

RESULT 23
 AZ528485
 LOCUS
 DEFINITION ENT528485 Entamoeba histolytica Sheared DNA linear genomic, genomic survey sequence.

AZ528485
 ENT528485
 DEFINITION 795 bp DNA linear GSS 03-NOV-2000
 Entamoeba histolytica Sheared DNA
 genomic, genomic survey sequence.

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,093
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-55

Query Match 100.0%; Score 2389; DB 3; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TTCTTACAGTTGGGACTGTATCAAGCTTAGAACCGTTAAAGGAAATAATCGTGTTCCTA	60
Db	1	TTCTTACAGTTGGGACTGTATCAAGCTTAGAACCGTTAAAGGAAATAATCGTGTTCCTA	60
Qy	61	TATAGATGGAACCAAGCGCGCAAAACCGGAAATTTGACTCCTGATGAGGTTAGCAA	120
Db	61	TATAGATGGAACCAAGCGCGCAAAACCGGAAATTTGACTCCTGATGAGGTTAGCAA	120
Qy	121	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTAC	180
Db	121	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTAC	180
Qy	181	TTCCATGCGGCAACCATATCAATATTCAATGTTAAGTTTCCTTATGACGCTATCATCAG	240
Db	181	TTCCATGCGGCAACCATATCAATATTCAATGTTAAGTTTCCTTATGACGCTATCATCAG	240
Qy	241	TGAAGATTACTCATGAAAGATCCAACTAAGCTAAAGATGAGGATATTGTTAAGA	300
Db	241	TGAAGATTACTCATGAAAGATCCAACTAAGCTAAAGATGAGGATATTGTTAAGA	300
Qy	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTACCTTAAGGATGC	360
Db	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTACCTTAAGGATGC	360
Qy	361	TGCCCGACGGATACCGTCCGTACAAAGAGGAATCAATGACAAACCAAGAGCATAG	420
Db	361	TGCCCGACGGATACCGTCCGTACAAAGAGGAATCAATGACAAACCAAGAGCATAG	420
Qy	421	TCAACATCGTCAAGGTGGAATCCCAAGAAACGATGGTGTGCTTGGCAGCGTTCGCA	480
Db	421	TCAACATCGTCAAGGTGGAATCCCAAGAAACGATGGTGTGCTTGGCAGCGTTCGCA	480

Qy	481	AGGACCTTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGTAC	540
Db	481	AGGACCTTACTACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGTAC	540
Qy	541	TGCTGATGCTTATATGTTTCTTCATCGAGATCATACCATATCATTCCTTAAGATGATT	600
Db	541	TGCTGATGCTTATATGTTTCTTCATCGAGATCATATACCATATCATTCCTTAAGATGATT	600
Qy	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAGCTTCTCTATCTTGGTCGAGGAAATCTGTCAA	660
Db	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAGCTTCTCTATCTTGGTCGAGGAAATCTGTCAA	660
Qy	661	TTCAAGAACCTATCGCCGCAAAATAGCGATAACATCTTCAAGAACCAACTGGGTACCTTC	720
Db	661	TTCAAGAACCTATCGCCGCAAAATAGCGATAACATCTTCAAGAACCAACTGGGTACCTTC	720
Qy	721	TGTAAGCAATCCAGGAACCTACAAATCTAACAGCAACCAAGCAACCACTAACAGTCA	780
Db	721	TGTAAGCAATCCAGGAACCTACAAATCTAACAGCAACCAAGCAACCACTAACAGTCA	780
Qy	781	AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTTACAACTGCCTTTGAG	840
Db	781	AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTTACAACTGCCTTTGAG	840
Qy	841	TCAACGACATGTAGAAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCAAC	900
Db	841	TCAACGACATGTAGAAATCTGATGCGCTTGTCTTTGATCCAGCACAATCAAGTCCAAC	900
Qy	901	AGCTAGAGGTTGCGAGTCCACCGAGATCATACCACTTCATCCCTTACTCTCAAAAT	960
Db	901	AGCTAGAGGTTGCGAGTCCACCGAGATCATACCACTTCATCCCTTACTCTCAAAAT	960
Qy	961	GTCTGAATTTGGAAGAACGAATCGTGTATTTTCCCTTCTGTTTATCGTTTCAAACTG	1020
Db	961	GTCTGAATTTGGAAGAACGAATCGTGTATTTTCCCTTCTGTTTATCGTTTCAAACTG	1020
Qy	1021	GGTACCAGATTCAGGCGCAACCAAGTCCACACCGACTCCGAACTAGTCCAGG	1080
Db	1021	GGTACCAGATTCAGGCGCAACCAAGTCCACACCGACTCCGAACTAGTCCAGG	1080
Qy	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCAAGTGGT	1140
Db	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTCAAGTGGT	1140
Qy	1141	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTCGTTATGTTTC	1200
Db	1141	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTCGTTATGTTTC	1200
Qy	1201	GAAAGATTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG	1260
Db	1201	GAAAGATTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAG	1260
Qy	1261	TGTTTACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCGTCAACCAAGAAATTTA	1320
Db	1261	TGTTTACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCGTCAACCAAGAAATTTA	1320
Qy	1321	TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGGTCG	1380
Db	1321	TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATAAGGGTCG	1380
Qy	1381	TAATCTGATTTCCAGCCTTAGCAAAATTTATAGACGCTTGAATGATGATCGACTAA	1440
Db	1381	TAATCTGATTTCCAGCCTTAGCAAAATTTATAGACGCTTGAATGATGATCGACTAA	1440
Qy	1441	TAAAGAAATTTGGTAGATGTTTATGGCAATCTTAGCACCANTTACCCATCCAGAGCG	1500
Db	1441	TAAAGAAATTTGGTAGATGTTTATGGCAATCTTAGCACCANTTACCCATCCAGAGCG	1500
Qy	1501	ACTTGGCAACCAAAATCTCAAAATGAGTATACATGAAAGCAAGTTGTTGTTGCTCAAT	1560
Db	1501	ACTTGGCAACCAAAATCTCAAAATGAGTATACATGAAAGCAAGTTGTTGTTGCTCAAT	1560
Qy	1561	AGCTGATAAGATATACAAAGTCAGATGGTTATATTTTGGATGAACATGATATAAATCAGTGA	1620

```

Db 1561 AGCTGATAAGTATACAACTGTCAGATGGTTTACATTTTGTGATGAATGATATATATCAGTGA 1620
Qy 1621 TGAAGGAGATGATATGTAACGCTCATATGCGCCATAGTCTAGTGTGGAAAGATAG 1680
Db 1621 TGAAGGAGATGATATGTAACGCTCATATGCGCCATAGTCTAGTGTGGAAAGATAG 1680
Qy 1681 CCTTCTCTGATAAGGAAAAAGTTGTCAGCTCAAGCTCATATCTAAAGAAAAAGGTATCTCTACC 1740
Db 1681 CCTTCTCTGATAAGGAAAAAGTTGTCAGCTCAAGCTCATATCTAAAGAAAAAGGTATCTCTACC 1740
Qy 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGCAGTATTTA 1800
Db 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTGCAGCAGTATTTA 1800
Qy 1801 CAATCGTGTGAAGGGGAAAAAGCAATCCACTCGTTCGACTTCATATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAGGGGAAAAAGCAATCCACTCGTTCGACTTCATATATATGTTGAGCA 1860
Qy 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTCTCTCATAGGATCATTTACCATATAT 1920
Db 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTCTCTCATAGGATCATTTACCATATAT 1920
Qy 1921 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGCTATACCTTGGAAAG 1980
Db 1921 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGCTATACCTTGGAAAG 1980
Qy 1981 TTTGTTTCGGGAGATTAAGTACTAGTAGAACCCTGACGAGCTCCACATTTAAATGA 2040
Db 1981 TTTGTTTCGGGAGATTAAGTACTAGTAGAACCCTGACGAGCTCCACATTTCTAATGA 2040
Qy 2041 TGGATGGGGCAATGCGAGTGAAGTGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
Db 2041 TGGATGGGGCAATGCGAGTGAAGTGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
Qy 2101 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Db 2101 TAAGAACTTCAAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
Qy 2161 TCAAGTAGAGCTGAAAAGTAGAGGCCCACTCAAGAGCAGAGATTTGCTTGGGAA 2220
Db 2161 TCAAGTAGAGCTGAAAAGTAGAGGCCCACTCAAGAGCAGAGATTTGCTTGGGAA 2220
Qy 2221 AGTAAGCGATCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTACGAAATAA 2280
Db 2221 AGTAAGCGATCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTACGAAATAA 2280
Qy 2281 TTTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2340
Db 2281 TTTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAGAGCAGAAAAATTTACTTGC 2340
Qy 2341 GTTGTAAAAGGAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389
Db 2341 GTTGTAAAAGGAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389

```

RESULT 2

```

US-09-536-784-55
; Sequence 55, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

```

```

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-536-784-55

Query Match 100.0%; Score 2388; DB 4; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTTCTTACGAGTTGGGACCTGTATCAAGCTAGAACCGTTAAGGAAATAATCGTGTTCCTA 60
Db 1 TTTCTTACGAGTTGGGACCTGTATCAAGCTAGAACCGTTAAGGAAATAATCGTGTTCCTA 60
Qy 61 TATAGATGGAACAAAGCGACGCGAAAAACCGAGAAATTTGACTCTGTAGAGTTAGCAA 120
Db 61 TATAGATGGAACAAAGCGACGCGAAAAACCGAGAAATTTGACTCTGTAGAGTTAGCAA 120
Qy 121 GGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAGGCGTATGTCA 180
Db 121 GGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAGGCGTATGTCA 180
Qy 181 TTCAATCGGACCACTATCATTTTCAATGTTAAGTTCCTTATGACGCTATCATCAG 240
Db 181 TTCAATCGGACCACTATCATTTTCAATGTTAAGTTCCTTATGACGCTATCATCAG 240
Qy 241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA 300
Db 241 TGAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGA 300
Qy 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 301 GGTCAAGGTTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Qy 361 TGCCCAACGCGATTAACGTCGTACAAAAGAGGAATCAATCGACAAAAACAGAGCATAG 420
Db 361 TGCCCAACGCGATTAACGTCGTACAAAAGAGGAATCAATCGACAAAAACAGAGCATAG 420
Qy 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGCTTGGCACCCTTGC 480
Db 421 TCAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGCTTGGCACCCTTGC 480
Qy 481 AGGAGCTTACTACAGATGATGTTTATTAATGCTTCTGATATCATAGAGGATAC 540
Db 481 AGGAGCTTACTACAGATGATGTTTATTAATGCTTCTGATATCATAGAGGATAC 540
Qy 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCAATACCATTTACATTCCTAAGATGATT 600
Db 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCAATACCATTTACATTCCTAAGATGATT 600
Qy 601 ATCAGTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGTCGAGGAAATCTGTCAA 660
Db 601 ATCAGTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGTCGAGGAAATCTGTCAA 660

```


661 TTCAAGAACTATCGCGGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 720
 661 TTCAAGAACTATCGCGGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC 720
 721 TGTAAAGCAATCCAGGAACTACAATACTACACAGCAACCAAGCAACCACTAACAGTCA 780
 721 TGTAAAGCAATCCAGGAACTACAATACTACACAGCAACCAAGCAACCACTAACAGTCA 780
 781 AGCAAGTCAAAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTAACAACCTGCTTTCAG 840
 781 AGCAAGTCAAAAGTAATGACATTCATAGTCTCTTTGAAACAGCTCTAACAACCTGCTTTCAG 840
 841 TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCAACCACTAACAGTCA 900
 841 TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCAACCACTAACAGTCA 900
 901 AGCTAGAGGTGTTCGAGTGCACACGAGAGATCAATACCACTTCACTCCCTTACTCTCAAT 960
 901 AGCTAGAGGTGTTCGAGTGCACACGAGAGATCAATACCACTTCACTCCCTTACTCTCAAT 960
 961 GTCGAAATTCGAAAGCAATCGCTCGTATATTTCCCTTCTGTTATCGTTCAAAACCAATTG 1020
 961 GTCGAAATTCGAAAGCAATCGCTCGTATATTTCCCTTCTGTTATCGTTCAAAACCAATTG 1020
 1021 GGTACAGATTCAGGCGCAACCAAGTCCACACGAGTCCCGAACCTAGTCCAGG 1080
 1021 GGTACAGATTCAGGCGCAACCAAGTCCACACGAGTCCCGAACCTAGTCCAGG 1080
 1081 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 1140
 1081 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGCTGGT 1140
 1141 ACAGAAATTCGAAAGCAATGATGATTCGAAAGAAAGGCAATCTCTGTTATGCTTTTC 1200
 1141 ACAGAAATTCGAAAGCAATGATGATTCGAAAGAAAGGCAATCTCTGTTATGCTTTTC 1200
 1201 GAAAGATTTACCACTCTCAAACTGTTAAATCTTGAAGCAAGTATCAAAACCAAGAG 1260
 1201 GAAAGATTTACCACTCTCAAACTGTTAAATCTTGAAGCAAGTATCAAAACCAAGAG 1260
 1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTA 1320
 1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTA 1320
 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAGGCTCG 1380
 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATAGGCTCG 1380
 1381 TAATCTGATTTCCAAAGCCTTAGACAAATATTTAGAACGCTTGAATGATGAATCGACTAA 1440
 1381 TAATCTGATTTCCAAAGCCTTAGACAAATATTTAGAACGCTTGAATGATGAATCGACTAA 1440
 1441 TAAAGAAAAATGCTAGATGATTTATTGGCAATCTTAGCAACCAATACCCATCCAGAGCG 1500
 1441 TAAAGAAAAATGCTAGATGATTTATTGGCAATCTTAGCAACCAATACCCATCCAGAGCG 1500
 1501 ACTTGGCAACCAATCTTCAATTTAGTATGATGATGAGCAAGTTCGTTATGCTCAATT 1560
 1501 ACTTGGCAACCAATCTTCAATTTAGTATGATGATGAGCAAGTTCGTTATGCTCAATT 1560
 1561 AGCTGATAAGTATCAAGCTCAGATGTTACATTTTGTGAAACATGATTAATCAAGTGA 1620
 1561 AGCTGATAAGTATCAAGCTCAGATGTTACATTTTGTGAAACATGATTAATCAAGTGA 1620
 1621 TGAAGGAGATGCATATGTAAGCCTCATATGCGGCCATAGTCACTGGATGGAAAGATAG 1680
 1621 TGAAGGAGATGCATATGTAAGCCTCATATGCGGCCATAGTCACTGGATGGAAAGATAG 1680
 1681 CCTTCTTGATAAGAAAAAGTTGAGCTCAAGCTTACTAAGAAAAAGTATCCCTACC 1740
 1681 CCTTCTTGATAAGAAAAAGTTGAGCTCAAGCTTACTAAGAAAAAGTATCCCTACC 1740

1741 TCCATCTCAGACGACAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
 1741 TCCATCTCAGACGACAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860
 1861 TACAGTGTAGGTTAAAGCGGTAATTTGATTTCTCTAAGGATCATTTACCATATAT 1920
 1861 TACAGTGTAGGTTAAAGCGGTAATTTGATTTCTCTAAGGATCATTTACCATATAT 1920
 1921 TAAATTTGCTTTGCTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
 1921 TAAATTTGCTTTGCTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980
 1981 TTTGTTTGGGACGATTAAGTACTACGTAGAACAACCTGACGAACGTCCTCAATTTCTAATGA 2040
 1981 TTTGTTTGGGACGATTAAGTACTACGTAGAACAACCTGACGAACGTCCTCAATTTCTAATGA 2040
 2041 TGGATCGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 2041 TGGATCGGGCAATGCGCAGTGAGCATGTTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
 2101 TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
 2101 TAAAGAACTTCAAGCGGATGAAGCCAGTGAAGGAAACACCTGCTGAGCCAGAAGTCCC 2160
 2161 TCAAGTAGAGACTGAAAAAGTAGAACGCCAACTCAAGAGCAAGAGTTTGTCTTGGAA 2220
 2161 TCAAGTAGAGACTGAAAAAGTAGAACGCCAACTCAAGAGCAAGAGTTTGTCTTGGAA 2220
 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA 2280
 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA 2280
 2281 TTTGACTCTTCAAAATTTATGATTAACATAGTATCATGGCAAGAGAGAAAAATTTACTTGC 2340
 2281 TTTGACTCTTCAAAATTTATGATTAACATAGTATCATGGCAAGAGAGAAAAATTTACTTGC 2340
 2341 GTTGTAAAGGAAGTAATCTTCACTCTGTAAGTGAAGAAAAAATAAAC 2389
 2341 GTTGTAAAGGAAGTAATCTTCACTCTGTAAGTGAAGAAAAAATAAAC 2389

RESULT 3
 US-09-468-656A-9
 ; Sequence 9, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; FILE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 US-09-468-656A-9
 Query Match 100.0%; Score 2389; DB 4; Length 2451;

		Best Local Similarity 100.0%; Pred. No. 0;				Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	TTCTTACGAGTGGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	60						
Db	60	TTCTTACGAGTGGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	119						
QY	61	TATAGATGGAAGAACAGCGACGCAAAAGGCGGAGAAATTTGACCTCGATGAGGTAGCAA	120						
Db	120	TATAGATGGAAGAACAGCGACGCAAAAGGCGGAGAAATTTGACCTCGATGAGGTAGCAA	179						
QY	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC	180						
Db	180	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC	239						
QY	181	TTACATGCGGACCACTATCAATTAATCAATGTAAGTTCCTTATGACGCTATCATAC	240						
Db	240	TTACATGCGGACCACTATCAATTAATCAATGTAAGTTCCTTATGACGCTATCATAC	299						
QY	241	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTAATGA	300						
Db	300	TGAAGAAATTAATCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTAATGA	359						
QY	301	GCTCAAGGTGGATATGTTATCAAGGTAGATGAAATAATCTATGTTTACCTTAAGGATGC	360						
Db	360	GCTCAAGGTGGATATGTTATCAAGGTAGATGAAATAATCTATGTTTACCTTAAGGATGC	419						
QY	361	TGCCACGCGGATAAGCTCGTCAAAAGAGGAAATCAATCGCAAAAGCAAGAGCATAG	420						
Db	420	TGCCACGCGGATAAGCTCGTCAAAAGAGGAAATCAATCGCAAAAGCAAGAGCATAG	479						
QY	421	TCAACATCTGAGGTGGAACTCCAAAGAAACGATGGTGTGCTGCGCTCGCGAGTTCGCA	480						
Db	480	TCAACATCTGAGGTGGAACTCCAAAGAAACGATGGTGTGCTGCGCTCGCGAGTTCGCA	539						
QY	481	AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC	540						
Db	540	AGGACGCTATATACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGGATAC	599						
QY	541	TGGTGATGCTTATATCGTTCCTCATGAGATCATTTACGATTTACATTCCTTAAGATGAT	600						
Db	600	TGGTGATGCTTATATCGTTCCTCATGAGATCATTTACGATTTACATTCCTTAAGATGAT	659						
QY	601	ATCAGTACGAGTGGCTGCTGCAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA	660						
Db	660	ATCAGTACGAGTGGCTGCTGCAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA	719						
QY	661	TTCAAGAACCTATCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC	720						
Db	720	TTCAAGAACCTATCGCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC	779						
QY	721	TGTAAGCAATCCAGGAATACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	780						
Db	780	TGTAAGCAATCCAGGAATACAAATACTAACACAGCAACACAGCAACACTAACAGTCA	839						
QY	781	AGCAAGTCAAAAGTAATGACATTTAGTCTCTTTGAAACAGCTCTACAACTCCCTTTGAG	840						
Db	840	AGCAAGTCAAAAGTAATGACATTTAGTCTCTTTGAAACAGCTCTACAACTCCCTTTGAG	899						
QY	841	TCAACGATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCACAAGTCGAAC	900						
Db	900	TCAACGATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCACAAGTCGAAC	959						
QY	901	AGCTAGAGTGTTCAGTCCACAGGAGATCATTTACGATTCCTTCTCTCAAT	960						
Db	960	AGCTAGAGTGTTCAGTCCACAGGAGATCATTTACGATTCCTTCTCTCAAT	1019						
QY	961	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATCCCTTCTGTTATCGTTCAAACTTG	1020						
Db	1020	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATCCCTTCTGTTATCGTTCAAACTTG	1079						
QY	1021	GGTACAGATTCAGGCGGAGAAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG	1080						

Db	1080	GGTACAGATTCAGGCGGAGAAACAAACAGTCCACAAACCGACTCCGAAACCTAGTCCAGG	1139
QY	1081	CCGCAACCTGCAACCAAACTCTTAAATAGACTCAAAATTTCTTTGTTAGTTCAGCTGGT	1140
Db	1140	CCGCAACCTGCAACCAAACTCTTAAATAGACTCAAAATTTCTTTGTTAGTTCAGCTGGT	1199
QY	1141	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTGTTATGCTTTGC	1200
Db	1200	ACGAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTGTTATGCTTTGC	1259
QY	1201	GAAAGATTTTACCACCTGAAACCTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAG	1260
Db	1260	GAAAGATTTTACCACCTGAAACCTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAG	1319
QY	1261	TGTTTCCACACCTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA	1320
Db	1320	TGTTTCCACACCTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA	1379
QY	1321	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG	1380
Db	1380	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCG	1439
QY	1381	TAATTCCTGATTTCCAAAGCCTTAGACAAATTTATAGAACCTTGAATGATCAATCGACTAA	1440
Db	1440	TAATTCCTGATTTCCAAAGCCTTAGACAAATTTATAGAACCTTGAATGATCAATCGACTAA	1499
QY	1441	TAAAGAAATTTGGTAGATGATTTATTTGCAATTCCTTAGCACCAATTAACCATCCAGGCG	1500
Db	1500	TAAAGAAATTTGGTAGATGATTTATTTGCAATTCCTTAGCACCAATTAACCATCCAGGCG	1559
QY	1501	ACTTGGCAACCAAAATTCCTCAATTTGAGTATATCTGAGACGAGTTGCTTATGCTCAAT	1560
Db	1560	ACTTGGCAACCAAAATTCCTCAATTTGAGTATATCTGAGACGAGTTGCTTATGCTCAAT	1619
QY	1561	AGCTGATAGTATACAACTGAGTGGTTACATTTTGTGATGAACATGATATATCACTGCA	1620
Db	1620	AGCTGATAGTATACAACTGAGTGGTTACATTTTGTGATGAACATGATATATCACTGCA	1679
QY	1621	TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCTAGATTGGAAAGATAG	1680
Db	1680	TGAAGAGATGATATGTAACGCTCATATGGGCCATAGTCTAGATTGGAAAGATAG	1739
QY	1681	CCCTTTCTGATAAGGAAAGTTGCGAGCTCAAGCCTATACCTAAAGAAAGGTTATCTACC	1740
Db	1740	CCCTTTCTGATAAGGAAAGTTGCGAGCTCAAGCCTATACCTAAAGAAAGGTTATCTACC	1799
QY	1741	TCCATCTCCAGACGAGATGTTTAAAGCAAACTCCAACTGAGATAGTGACAGAGTATTTA	1800
Db	1800	TCCATCTCCAGACGAGATGTTTAAAGCAAACTCCAACTGAGATAGTGACAGAGTATTTA	1859
QY	1801	CAATCGTGTGAAAGGGGAAACGAAATTCCTGCTCGACTTCATATATGTTGAGCA	1860
Db	1860	CAATCGTGTGAAAGGGGAAACGAAATTCCTGCTCGACTTCATATATGTTGAGCA	1919
QY	1861	TACAGTTCAGGTTTAAAGCGTTTATTTGATTTTCTCATAGGATCATTTACCATATAT	1920
Db	1920	TACAGTTCAGGTTTAAAGCGTTTATTTGATTTTCTCATAGGATCATTTACCATATAT	1979
QY	1921	TAAATTTGCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGCTATACCTTGAAGA	1980
Db	1980	TAAATTTGCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGCTATACCTTGAAGA	2039
QY	1981	TTTGTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA	2040
Db	2040	TTTGTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAAGCTCCACATTTCTAATGA	2099
QY	2041	TGGATGGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAAACACACAGTGAAGATCCAAA	2100
Db	2100	TGGATGGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAAACACACAGTGAAGATCCAAA	2159
QY	2101	TAAAGATTTCAAGCGGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	2160	TAAAGATTTCAAGCGGATGAAGCCAGTGAAGAAACACCTGCTGAGCCAGAGTCCC	2219

QY	2161	TCAAGTAGAGCTGAAAGTAGAAGCCCACTCAAGAAGCAGAGTTTGGCTTGGAA	2220	Db	3173	GGCTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTAC	3232
Db	2220	TCAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAAGCAGAGTTTGGCTTGGAA	2279	QY	181	TTCAATATGGGACCACTATCATTTATACATGTTAGGTTCTTATGACGCTATCATCAG	240
QY	2221	AGTAAAGGATCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACGAAATAA	2280	Db	3233	TTCAATATGGGACCACTATCATTTATACATGTTAGGTTCTTATGACGCTATCATCAG	3292
Db	2280	AGTAAAGGATCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACGAAATAA	2339	QY	241	TGAAGAAATTAATCATGAAGATCCAACTATAAGCTAAAAAGATGAGGATATGTTAATGA	300
QY	2281	TTTGACTCTTCAAATTTAGTAACTAATAGTATCATGCGAGAGCAGCAAAAATTTACTTGC	2340	Db	3293	TGAAGAAATTAATCATGAAGATCCAACTATAAGCTAAAAAGATGAGGATATGTTAATGA	3352
Db	2340	TTTGACTCTTCAAATTTAGTAACTAATAGTATCATGCGAGAGCAGCAAAAATTTACTTGC	2399	QY	301	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
QY	2341	GTGTTAAAGGAAGTAACTCTTCACTGTAAAGTAAAGGAAAAATAAAC	2389	Db	3353	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	3412
Db	2400	GTGTTAAAGGAAGTAACTCTTCACTGTAAAGTAAAGGAAAAATAAAC	2448	QY	361	TGCCAGCGGATAACTGCTCGTACAAAGAGGAATCAATCGACAAATAAAGAGCATAG	420
RESULT 4							
US-08-961-527-94							
; Sequence 94, Application US/08961527							
; Patent No. 6420135							
; GENERAL INFORMATION:							
; APPLICANT: Charles Kunsch							
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences							
; NUMBER OF SEQUENCES: 391							
; CORRESPONDENCE ADDRESS:							
; ADDRESSEE: Human Genome Sciences, Inc.							
; STREET: 9410 Key West Avenue							
; CITY: Rockville							
; STATE: Maryland							
; COUNTRY: USA							
; ZIP: 20850							
; COMPUTER READABLE FORM:							
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage							
; COMPUTER: HP Vectra 486/33							
; OPERATING SYSTEM: MSDOS version 6.2							
; SOFTWARE: ASCII Text							
; CURRENT APPLICATION DATA:							
; APPLICATION NUMBER: US/08/961,527							
; FILING DATE:							
; CLASSIFICATION: 424							
; PRIOR APPLICATION DATA:							
; APPLICATION NUMBER:							
; FILING DATE:							
; ATTORNEY/AGENT INFORMATION:							
; NAME: Brookes, A. Anders							
; REGISTRATION NUMBER: 36,373							
; REFERENCE/DOCKET NUMBER: PB340P1							
; TELECOMMUNICATION INFORMATION:							
; TELEPHONE: (301) 309-8504							
; TELEFAX: (301) 309-8512							
; INFORMATION FOR SEQ ID NO: 94:							
; SEQUENCE CHARACTERISTICS:							
; LENGTH: 8195 base pairs							
; TYPE: nucleic acid							
; STRANDEDNESS: double							
; TOPOLOGY: linear							
US-08-961-527-94							
Query Match							
Best Local Similarity 100.0%; Score 2388; DB 4; Length 8195;							
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;							
QY	1	TTCTTACGAGTTGGAGCTGTATCAAGCTTAGACGGTTAAGGAAATATCGTTTCTTA	60	QY	1081	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT	1140
Db	3053	TTCTTACGAGTTGGAGCTGTATCAAGCTTAGACGGTTAAGGAAATATCGTTTCTTA	3112	Db	4133	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGGT	4192
QY	61	TATAGATGAAACAAAGCAGCCCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120	QY	1141	ACGAAAAGTTGGGGAAGGATATGTTTCGAAGAAAAGGCACTCTCTGTTATGTTTTC	1200
Db	3113	TATAGATGAAACAAAGCAGCCCAAAACCGAGAAATTTGACTCTCTGATGAGTTAGCAA	3172	Db	4193	ACGAAAAGTTGGGGAAGGATATGTTTCGAAGAAAAGGCACTCTCTGTTATGTTTTC	4252
QY	121	CGGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC	180	QY	1201	GAAAGATTTACCATCTGAAACTGTTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG	1260

Db 4253 GAAAGATTTACCATCTGAAACCTGTTAAAACTCTTGAAGCAAGTTATCAAAACAAGAGAG 4312

Qy 1261 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTGCTGACCAAGAAATTTTA 1320

Db 4313 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTGCTGACCAAGAAATTTTA 4372

Qy 1321 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCGAAATTAAGGGTCG 1380

Db 4373 TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCGAAATTAAGGGTCG 4432

Qy 1381 TAATTCGATTTCCAGCCCTTAGACAAATTTATGAACGGCTTGAATGATGAATGACTAA 1440

Db 4433 TAATTCGATTTCCAGCCCTTAGACAAATTTATGAACGGCTTGAATGATGAATGACTAA 4492

Qy 1441 TAAAGAAAAATTCGTAGATGATTTATTTGGCATTTCTAGACCAATACCCATCCAGAGG 1500

Db 4493 TAAAGAAAAATTCGTAGATGATTTATTTGGCATTTCTAGACCAATACCCATCCAGAGG 4552

Qy 1501 ACTTGGCAAAACCAATTTCTCAAAATTCAGTATATCTGAAGACGAAGTTTCGTATTCCTCAAT 1560

Db 4553 ACTTGGCAAAACCAATTTCTCAAAATTCAGTATATCTGAAGACGAAGTTTCGTATTCCTCAAT 4612

Qy 1561 AGCTGATAGTATACACGCTCAGATGTTTACATTTTGTGATGAACATGATATATCAGTGA 1620

Db 4613 AGCTGATAGTATACACGCTCAGATGTTTACATTTTGTGATGAACATGATATATCAGTGA 4672

Qy 1621 TGAAGGAGATGCTATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680

Db 4673 TGAAGGAGATGCTATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 4732

Qy 1681 CTTTCTGTAAGGAAAAAGTTTCAGCTCAAGCTATATCTAAAGAAAAAGGTTATCTTACC 1740

Db 4733 CTTTCTGTAAGGAAAAAGTTTCAGCTCAAGCTATATCTAAAGAAAAAGGTTATCTTACC 4792

Qy 1741 TCCATCTCAGACGCGAGATGTTAAGCAATCCAACTGGAGATAGTGCAGCACTATTTA 1800

Db 4793 TCCATCTCAGACGCGAGATGTTAAGCAATCCAACTGGAGATAGTGCAGCACTATTTA 4852

Qy 1801 CAATCGTGTGAAAGGGAAAAAGCAATTCACCTCGTTCCACTTCCATATATGTTGTCAGCA 1860

Db 4853 CAATCGTGTGAAAGGGAAAAAGCAATTCACCTCGTTCCACTTCCATATATGTTGTCAGCA 4912

Qy 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTAATTCCTCATAGGATCATTTACCATAATAT 1920

Db 4913 TACAGTTGAGGTTAAAAACGGTAATTTGATTTAATTCCTCATAGGATCATTTACCATAATAT 4972

Qy 1921 TAAATTTGCTTGTGATGATCACAATACAAAGCTCCAAATGGGTATACCTTGAAGA 1980

Db 4973 TAAATTTGCTTGTGATGATCACAATACAAAGCTCCAAATGGGTATACCTTGAAGA 5032

Qy 1981 TTTGTTTGGCAGATTTAAGTACTACGTAGAACACCTGACGACGTCACATTTCTAATGA 2040

Db 5033 TTTGTTTGGCAGATTTAAGTACTACGTAGAACACCTGACGACGTCACATTTCTAATGA 5092

Qy 2041 TGGATGGGCAATGTCAGTGAGCATGTTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100

Db 5093 TGGATGGGCAATGTCAGTGAGCATGTTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 5152

Qy 2101 TAAGAACTTCAAGCGGATGAGAGCCAGTGAAGAAACACCTGTCAGCCAGAGTCCC 2160

Db 5153 TAAGAACTTCAAGCGGATGAGAGCCAGTGAAGAAACACCTGTCAGCCAGAGTCCC 5212

Qy 2161 TCAAGTAGAGATGTAAGAAAGTGAAGCCCACTCAAGAAAGCAGAGTTTTCCTTGGGAA 2220

Db 5213 TCAAGTAGAGATGTAAGAAAGTGAAGCCCACTCAAGAAAGCAGAGTTTTCCTTGGGAA 5272

Qy 2221 AGTAACGGAATTCAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA 2280

Db 5273 AGTAACGGAATTCAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGTTTACGAAATAA 5332

Qy 2281 TTTGACTCTTCAAAATTCGATTAACATAGTATCATGGCAGAGCAAGAAATTAATCTGC 2340

Db 5333 TTTGACTCTTCAAAATTCGATTAACATAGTATCATGGCAGAGCAAGAAATTAATCTGC 5392

Qy 2341 GTTGTAAAAAGGAAGTAACTCTTCATCTGTAAGTAAGAAAAATAAAC 2389

Db 5393 GTTGTAAAAAGGAAGTAACTCTTCATCTGTAAGTAAGAAAAATAAAC 5441

RESULT 5

US-09-468-656A-11

; Sequence 11, Application US/09468656A

; Patent No. 6582706

; GENERAL INFORMATION:

; APPLICANT: Johnson, Leslie S.

; APPLICANT: Adamou, John E.

; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural

; TITLE OF INVENTION: Motifs

; FILE REFERENCE: 469201-444

; CURRENT APPLICATION NUMBER: US/09/468,656A

; CURRENT FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/113,048

; PRIOR FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11

; LENGTH: 2531

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-468-656A-11

Query Match 41.5%; Score 991; DB 4; Length 2531;

Best Local Similarity 65.7%; Pred. No. 1.5e-264;

Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57

Db 60 TTCTTATAGCTTGGAGCTTACCAAGCTGGTCAAGATTAAGAAGAGCTTAATCGAGTTGC 119

Qy 58 CTATATAGATGGAACAACAGCGACGCAAAAACGGAGAAATTTGACTCTCGATGAGTTAG 117

Db 120 TTATATAGATGTTGATCAGGCTGGTCAAAAGGCGAGAAAACCTTGACACAGATGAAGTCAG 179

Qy 118 CAAGCTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGACAAGGCTATGT 177

Db 180 TGAAGGGAGGGATCAACCGCGAACAAATGTTTATCAAGATTAACGATCAAGGTTATGT 239

Qy 178 CACTTCACATGGCGACCACTATCATTTATCAATGTAAGGTTCCCTTATGACGCTATCAT 237

Db 240 GACCTCTCATGGAGACCAATTATCATTTACTATTAATGSCAAGTTCCCTTATGATGCCATCAT 299

Qy 238 CAGTGAAGAAATTAATCATGAAAGATCCAAACTATAGCTTAAAGATGAGGATATGTTAA 297

Db 300 CAGTGAAGAGCTCCCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTTGCAA 359

Qy 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATCGAAATACTATGTTTACCTTAAGCA 357

Db 360 TGAATCAAGGGTGGTATGTCATTAAGGTAAACGGTAAATTAATTAATGTTTACCTTAAGCA 419

Qy 358 TGCTGCCCAACCGGATAACGTCCTGTCACAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417

Db 420 TGCRCCTCATCGGATAATATTTCGGACAAAAGAGAGATTAACCGTCAGAGCAGAAAGC 479

Qy 418 TAGTCAACATCGTGAAGTGGAACTCCCAAGAAACGATGGTGTGCTGTGGCAGCTTC 477

Db 480 CAGTCAATATCAT-----AACTCAAGAGCAGATAATGCTGTTGCTGACGACGAGC 530

Qy 478 GCAAGGAGCTTATACAGATGATGTTTATCTTTTATGCTTCTGATATCATAGAGA 537

Db 531 CCAAGGAGCTTATACAGGATGATGGGTATATCTTCAATGCATCATGATCATTAATGAGA 590

Qy 538 TACTGGTATGCTTATATCGTTTCTTCATGAGATCAATACCATTAACATTCTTCAAGAAATGA 597

Db 591 CACGGGTGATGCTTATATCGTTTCTTCACGGCGACCAATTAACATTACATTCCTAAGAAATGA 650

QY 598 GTTATCAGCTAGGAGTTGGCTGCTCAGAGAGCTTCTTATCTGCTCGAGAAATCTGTC 657
 Db 651 GTTATCAGCTAGGAGTTAGCTGCTGAGAGAGCTTATGGAATGG----- 695
 QY 658 AAATTCAAGAACCTTATCGCGACAAATAGCGATAACACTTCAAGAACAAACTGGTACC 717
 Db 696 -GAAGCAGGATCTCGTCTCTTCTCAAGTTCTAGTTATATGAAATCCAGCTCAACCAA 754
 QY 718 TTCTGTAAAGCAATCCAGGAATACTAAATATACTAAACAGCAACACAGCACTAACAG 777
 Db 755 GATTGTGAGAGAACCAATCT-----GACTGTCTCACTCAACTTA 794
 QY 778 TCAAGCAAGTCAAGTAATGACATGATAGTCTCTTGAAACAGCTCTACAAATGCTCTTT 837
 Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTACGTGAATGTATGCTTAACCTT 854
 QY 838 GAGTCAACAGCATGTGAATCTGATGGCTTGTCTTTGATCCAGCACAAATCAAGTGG 897
 Db 855 ATCAGACGCCATGTGGAATCTGATGGCTTTATTTTCGACCGCAAAATCAAGTGG 914
 QY 898 AACAGCTAGAGGTGTTGGAGTCCACACAGGAGATCATACCACCTCATCCCTTACTCTCA 957
 Db 915 AACGCCAGAGGTGTAGTGTCTCTCATGTGTAACCATTAACCATTTATCCCTTTATGAACA 974
 QY 958 AATGCTGAATTGGAAGAACGAATCGCTGCTGATTAATTTCCCTTATCGTTATCGTTCAAAACA 1017
 Db 975 AATGCTGAATTGGAAGAACGAATGCTGCTGATTAATTTCCCTTATCGTTATCGTTCAAAACA 1034
 QY 1018 TTGGGTACAGATTTCAAGGCCAGAACCAACCAAGTCCACACCGACTCGGGAACCTAGTCC 1077
 Db 1035 TTGGGTACAGATTTCAAGGCCAGAACCAACCAAGTCCACACCGACTCGGGAACCTAGTCC 1094
 QY 1078 AGGCCGCAACTGCAACCAATCTTAATAGACTCAAAATCTTCTTTGTTAGTCACT 1137
 Db 1095 AAGTCCGCAACAGCTTCAAGCAATCCAAATGATGGGAA-----ATTGGTCAAGAGC 1148
 QY 1138 GGTACGAAAGTGGGAGAGATGTATTCGAGAAAGGCACTCTCTGTTATGTTCT 1197
 Db 1149 TGTTGAAAGTAGGCGATGTTATGTTCTTTGAGAGAAATGAGTTTCTGTTATATCCC 1208
 QY 1198 TCGAAAGATTTACCATCTGAATCTTAAATCTTGAAGCAAGTTATCAAAACAAGA 1257
 Db 1209 AGCCAAGATCTTTCAAGCAAGAACAGCAGCAGGCAATGATGCAAACTGGCCAGCAGGA 1268
 QY 1258 GAGTGTTCACACATTTAACTGCTTAAAGAAAGAAATGTTCTCTGCTGACCCAGAAAT 1317
 Db 1269 AAGTTTATCTATAAGTAGGAACTTAAGAAACTGACCTCCCATCTAGTGAATCGAAAT 1328
 QY 1318 TTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTTGTTGNAATAAGGG 1377
 Db 1329 TTACATTAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACTTTGATAATAAGG 1388
 QY 1378 TCCTAATTTCTGATTTCCAGGCTTATAGCAAAATTTAGAACCTTGAATGAATCGAC 1437
 Db 1389 TCGCAAGTTGATTTTGAGGCTTTGATTAACCTGTTGGAACGACTCAAGGATGTCCTAAG 1448
 QY 1438 TAATAAGAAATTTGGTAGATGATTTATGTCATTTCTAGCACCAATTAACCCATCAGA 1497
 Db 1449 TGATAAGTCAAGTAGTGAAGATATCTTCCCTTTCTAGTCCGATTCGTCATCCAGA 1508
 QY 1498 GCGACTTGGCAACCAAAATCTCAAAATGAGTATATCTAGAGCAGAGTTGCTGCTCA 1557
 Db 1509 AGCTTTAGGAAACCAATGCGCAATTAACCTACACTGATGATGATTAAGTAGTCCAA 1568
 QY 1558 ATTACTGATTAAGTATACAGCTCAGATGTTTACATTTTGTATGACATGATATATCAG 1617
 Db 1569 GTTGGCAGGCAAGTACACAGCAGAGACGCTTATCTTTGATCTCTCGTGAATACACAG 1628
 QY 1618 TCATGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGAATGGAAGAGA 1677
 Db 1629 TGATGAGGGGATGCTTATGTAACCTCAATATGATGATGATGATGATGATGATGATGAT 1688
 QY 1678 TAGCCCTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTAAGAAAAAGGATTCCT 1737

Db 1689 TAGTTGTCTGAAGCTGAGAGAGCGGCGAGCCCGCTTATGCTTAAAGAGAAAGTTGAC 1748
 QY 1738 ACCTCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTAT 1797
 Db 1749 CCCTCTCTGACAGACCATCAGGATTCAGGAATATCTGAGCAAAAGGACAGAGCTAT 1808
 QY 1798 TTACAATCTGTGTGAAGGGGAAAAACGAATTTCCACTGTTTCGACTTCCATATATGGTTGA 1857
 Db 1809 CTACACCCGCTGAGAGCAGCTAAGAGGTGCGCACTTGAATGCTTATGCTTACATCTCA 1868
 QY 1858 GCATACAGTTGAGGTTAAACCGGTATTTGATTTATTCCTCATAGAGATCATTAACCAATA 1917
 Db 1869 ATATCTGTAGAAGTCAAAACCGTGTATTAATCATACCTCATTAATGATTAACCAATA 1928
 QY 1918 TATTAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGA 1977
 Db 1929 CATCAAAATTTGAGTGTGTTGACGAAGCTTATGAGCAGCTTAAGGGGTATCTCTGA 1988
 QY 1978 AGATTTGTTTGGCAGCAATTAAAGTACTAGTAGAACACCCCTGAGAACAGCTCCACATTTCTAA 2037
 Db 1989 GGATCTTTTGGCAGCTGCTCAAGTACTATGTGAAACATCCAAACGAGCTCGCATTCAGA 2048
 QY 2038 TGATGATGGGGCAATCCAGTGAAGTGTGTTAGGCAAGAAAGACCCACAGTGAAGTCC 2097
 Db 2049 TAATGTTTGTGTAACGCTAGCGACCATGTTCAAAAGAAACAAAAATGGTCAAGCTGATC 2108
 QY 2098 AAATAAGCACTTCAAAAGCGGATGA----- 2121
 Db 2109 CAATCAAAACGAAAAACCAAGCAGGAGGAAACCTTCAGACAGAAAAACCTGAGGAGAAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCCTCGAGAGAGAAAAACCGAAAGCGAGAAACACAGAGTCTCCAAAAACCAACAGAGAAC 2228
 QY 2131 AGAGAAAAACCTCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAGTAGAGCCCA 2190
 Db 2229 AGAAGATTCAGCAGAGGAATCAGAAAGAACCTTCAGTGCAGACTGAAAGGTTGAAGAAA 2288
 QY 2191 ACTCAAGAAAGCAGAAAGTTTGTCTTGGAAAGTAACGGAATCTAGTCTGAAAGCCAAATGC 2250
 Db 2289 ACTGAGAGGCTGAGAGATTTACTTTGNAAAATCCAGATCCAATTAAGTCCAATGC 2348
 QY 2251 AACGAAACTCTAGTGTGTTTACGAAATAATTTGACTCTTCAAAATATTGATTAACATAG 2310
 Db 2349 CAAAGAGACTCTCAGAGATTAATAATTTACTATTGTCACCCAGGACCAACAATAC 2408
 QY 2311 TATCATGCAAGCAGAGAAAAATTTACTTGCCTTGTAAAAAGGAAGTAA 2358
 Db 2409 TATTATGCAAGAGCTGAAAAAACTATTGCTTTTATTAAGGAGAGTAA 2456

RESULT 6
 US-09-468-656A-5
 ; Sequence 5, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: 1999-12-02
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae

US-09-468-656A-5

Query Match 41.4%; Score 990; DB 4; Length 2531;
 Best Local Similarity 68.1%; Pred. No. 2.8e-264;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 1 TTCTTACGAGTGGGACTGTATCAAGCTAGAAACGGTTTAAGGAAAA---TAATCGTGTTC 57
 DB 60 TTCTTATGAACCTTGGTGGTCAACCAAGCTGGTCAAGGTTAAGAAAGAGTCTAATCGAGTTTC 119
 QY 58 CTATATAGATGGAAGAACACAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAG 117
 DB 120 TTATATAGATGGTGTATCAGGCTGGTCAAAAGGAGGAGAAATCTTGACACCAAGATGAAGTCAG 179
 QY 118 CAAGCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGT 177
 DB 180 TAAAGGGAGGGGATCAACCGCCGAACAAATCGTCATCAAGATACAGGATCAAGGTTATGT 239
 QY 178 CACTTCACATGGCGACCACTATCATTAATTAACAATGGTAAGGTTCCCTTATGACGCTATCAT 237
 DB 240 GACCTCTCATGGAGACCAATATCATTAATTAATGGCAAGGTCCTTATGATGCCATCAT 299
 QY 238 CAGTGAAGAAATTAATCATGAAGATCAAACTATAGCTTAAAGATGAGAGATATGTTAA 297
 DB 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATATCAAGTTGAAGGATTCAGACATTTGTCAA 359
 QY 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 357
 DB 360 TGAATCAAGGTGGTATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA 419
 QY 358 TGCTGCCACCGGATAACGTCCTGATCAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
 DB 420 TGCAGCTCATCGGATAATATTCGGACAAAGAGAGATTAACGTCAGAGCAGGAACA 479
 QY 418 TAGTCAACATCTGGAAGGTGGAATCAAGAAACGATGGTGGTGGTCCCTTGGCAAGTTTC 477
 DB 480 CAGTCAATATCAAGGGGGTGGTCT-----AACGATCAAGCAGTAGTGGCAGCCAGAGC 533
 QY 478 GCAAGGACGCTATACPAACAGATGATGGTTATATCTTTAATGCTTCATGATATCATAGGA 537
 DB 534 CCAAGGACGCTATACCAAGGATGATGGTTATATCTTCAATGCACTGATATCATGAGGA 593
 QY 538 TACTGGTATGCTTATATCGTCTCATGGAGATCAATACCAATTAATTCCTAAGAAATGA 597
 DB 594 CACGGGTGATGCTTATATCGTCTCATCGGACCAATACCAATTAATTCCTAAGAAATGA 653
 QY 598 GTTATCAGCTAGGAGTTGGCTGCTGCAAGGCTTCCTATCTGGTTCGAGGAAATCTGTC 657
 DB 654 GTTATCAGCTAGGAGTTAGCTGCTGCAAGGCT----- 688
 QY 658 AAATTCAGAACTATCGCCGACAAATAGCGATACACTTCAAGAAACAACTGGGTACC 717
 DB 689 -----ATTGGAATGGAAGCAGGGATCTGCTCTCTTCAAGTTCTAGTTTAAATGC 740
 QY 718 TTCTGTGAAGCAATCCAGGAACCTACAAATACCAAGCAACCAAGCAACCACTAACAG 777
 DB 741 AAATCCAGCTCAACCAAGATTCAGAGAACCAACCAATCTGACTGCTCACTCAACTATCA 800
 QY 778 TCAAGCAAGTCAAGTAAATGACATTAATGATCTCTTTGAAACAGCTCTACAAATCGCTTT 837
 DB 801 TCA---AAATCAAGGGGAAAAAATTTCAAGGCTTTTACGTTGAATGTTATGCTTAAACCCCTT 857
 QY 838 GAGTCAACGACATGTAGAATCTGATGGCTGTGTTTGTATCCAGACCAATCACAGTGC 897
 DB 858 ATCAGAACCCATGTGGATCTGATGGCCCTTATTTTCGACCCAGCGCAATCACAGTGC 917
 QY 898 AACAGCTAGAGGTGTGAGTGGCAACAGGAGATCAATACCACTTTCATCCCTTACTCTCA 957
 DB 918 AACCGCCAGAGGTAGTGTCCCTCATGTTAAACCAATACCACTTATCCCTTATGAACA 977
 QY 958 AATGCTGAAATTTGGAAGAACGAATCGCTGATATATTCCTCTTGGTATCGTTTCAAAACA 1017
 DB 978 AATGCTGAAATTTGGAAGAACGAATTTGCTGATATATTCCTCTTGGTATCGTTTCAAAACA 1037

QY 1018 TTGGTACCAGATTCAAGGCCAGAACCAAGTCCACCAAGCTCCGAACTCGTAGTCC 1077
 DB 1038 TTGGTACCAGATTCAAGGCCAGAACCAAGTCCACCAAGCTCCGAACTCGTAGTCC 1097
 QY 1078 AGGCCGCAACCTCGACCAATCTTAAATPAGACTCAAAATCTTC-----TTT 1125
 DB 1098 AAGTCGCAACCTCGACCAATCTTCAACAGCTCCAAAGCTCCAAATTCGATGAGAAAT 1157
 QY 1126 GGTATGCTAGCTGGTACGAAAGCTTGGGAGAGATATGTATTCGAAAGAAAGGCGATCTC 1185
 DB 1158 GGTCAAAAGAGCTGTTTCGAAAGCTAGGCGATGTTATGCTTTTGGAGAGAAATCGAGTTTC 1217
 QY 1186 TCGTTATGCTTTTGGCAAGATTTACATCTGAACTGTTTAAAAATCTTGAAGAGCAAGTT 1245
 DB 1218 TCGTTATATATCCAGCCCAAGGATCTTTTTCAGCAGAAACAGCAGCAGGATTTGATAGCAAACT 1277
 QY 1246 ATCAAAACAAGAGAGTGTTCACACACTTTTAACTGCTTAAAAAGAAATGTTCTCTCTCG 1305
 DB 1278 GGCAGAGCAGGAAGTTTATCTCATAGCTAGAGCTTAAAGAACTGACCTCCATCTAG 1337
 QY 1306 TGACCAAGAAATTTTATGATAAAGCATATAATCTGTTAACTAGGCTCATAAAGCCTTGT 1365
 DB 1338 TGATCGAGAATTTTCAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1397
 QY 1366 TGAANAATAAGGTCGTAATCTGATTTTCAAGCTTAGACAAATTTAGAACGCTTGAA 1425
 DB 1398 TGATAAATAAGGTCGCAAGTGTGATTTTGGAGCTTTGGAACCTGTTGGAACGACTCA 1457
 QY 1426 TGATGAATCGACTAATAAAGAAATTTGGTAGATGATTTTATTCGCATTTCTGACCACT 1485
 DB 1458 GGAATGCTCCAGTGATTAAGTCAAGTTAGTGGATGATATCTTGGCTTTCTAGCTCCGAT 1517
 QY 1486 TACCCATCCAGAGCGACTTGGCAACCAAAATCTCAAAATGGAGTATATCTGAAAGCAAGT 1545
 DB 1518 TCGTATCCAGAACGTTTAGGAAACCAAAATGCGCAATTTACCTACACTGATGATGAGAT 1577
 QY 1546 TCGTATGCTCAATAGCTGATAAGTATACAGCTCAGATGTTTACATTTTGTATGAACA 1605
 DB 1578 TCAAGTAGCCTAAGTTGGCAGGCAAGTACACACAGAAACGCTTATCTTTGATCCTCG 1637
 QY 1606 TGATTAATCACTGATGAAGGAGATGCATATGTAACGCTCATATGGGCTATAGTCACTG 1665
 DB 1638 TGATTAACCACTGATGAGGCGGATGCTTATGTAATCTCCACATATGACCATAGCCACTG 1697
 QY 1666 GATTTGAAAGATAGCTTTCTGATAAGGAAAGTTGAGCTCAAGCTTATCTATAAGA 1725
 DB 1698 GATTAATAAAGATAGTTTGTCTGAAAGCTGAGAGAGCGGAGCCAGGCTTATGCTAAAGA 1757
 QY 1726 AAAAGGTATCTACCTCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAGATAG 1785
 DB 1758 GAAAGTTTGAACCTCTCTTCGACAGACCATCAGGATTCAGGAAATCTGAGGCAAAAG 1817
 QY 1786 TGCAGCAGCTATTTACAATCTGTGAAAGGGGAAAAACGAAATTCACCTCGTTCGACTCC 1845
 DB 1818 ASCAGAGCTATCTACACCGCTGAAAGCAGCTAAGAGGTGCCACTTGATCGTATGCC 1877
 QY 1846 ATATATGTTGAGCATACAGTTGAGTTAAACCGGTAAATTTGATTTATTTCTCATAAGA 1905
 DB 1878 TTACAATCTTCAATATCTGTAGAAAGTCAAAAACGGTAGTTTAAATCATACCTCATATGA 1937
 QY 1906 TCATTTACCAATATTAATTTGCTTGGTGGTGGATGATCACATACAAAGCTCCAATGG 1965
 DB 1938 CCAATACCAATCAATCAAAATTTGAGTGGTTTGAAGAAGCCCTTTATGAGGCACTTAAGG 1997
 QY 1966 CTATACCTTGAAGATTTGTTTGGCAGCATTAAGTACTAGTAGAACACCTCGAGCAAG 2025
 DB 1998 GTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTTCGAACATCCAAAGCAAG 2057
 QY 2026 TCCACATTTCTAATGATGGAGGGCAATGCCAGTGGAGATGTTTGGCAAGAAAGACCA 2085
 DB 2058 TCCGCAATTCAGATAATGTTTGGTAAAGCTAGCGACCATGTTTCGTAATAAATAAGGTAGA 2117


```
QY 1366 TGAATAAAGGTCCTAATCTGATTTCCAGACCTTAGACAAATATTAGAACGCTTCAA 1425
Db 1339 TGATAATAAGGTCGACAGTTGATTTGAGCTTTGGATTAACCTGTTGGAAACGACTCAA 1398
QY 1426 TGATGAATCGACTAATAAGAAATAATGGTAGATGATTTATTTGGCAATTCCTAGACCAAT 1485
Db 1399 GGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATTCTTGCCCTTTAGCTCCGAT 1458
QY 1486 TACCATCCAGAGGACTTCGCAACCAATTTCTCAATTCAGTATGATGATGAGATGAGAT 1545
Db 1459 TCGTCATCCAGAGGTTTAGGAAACCAATGCGCAAAATTAACCTACACTGATGATGAGAT 1518
QY 1546 TCGTATTGCTCAATTAGCTGATAGTATACAACTGAGATGTTTACATTTTGTATGAACA 1605
Db 1519 TCAAGTAGCCAAGTTGCGAGCAAGTATACACAGAGAGCGTTTATATCTTTGATCTCG 1578
QY 1606 TGATATAATCAGTGATGAAGAGATGATATGTAACGCTCATATGGGCGCATAGTCACTG 1665
Db 1579 TGATATAACAGTGATGAGGGGATGCTATGTAACCTCCATATGACCCATAGCCACTG 1638
QY 1666 GATTGGAAGAGATAGCTTTCTGATAAGGAAAGTTGCGAGTCAAGCTCAAGCTTACTAAGA 1725
Db 1639 GATTAAAGAGATAGTTTGTCTGAAGCTGAGAGCGGCGAGCCGAGGCTTTATGCTAAGA 1698
QY 1726 AAAAGGTATCTACCTCCATCTCCAGACGAGATGTTAAAGCAAACTCCAACCTGAGATAG 1785
Db 1699 GAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATTAAGGCAAAAGG 1758
QY 1786 TGCAGCAGCTATTACAACTGCTGAAAGGGGAAACGAAATTCACCTGCTCGACTTCC 1845
Db 1759 AGCAGAGGCTATCTCAACCGCTGAAAGCAGCTAAGAAAGTGCCTTATGCTATGCC 1818
QY 1846 ATATATGTTTGACATACAGTTGAGTTTAAAGCGGTAAATTTGATTTCTCTATGAAGA 1905
Db 1819 TTCAATCTTCATATACCTGTAAGTCAAAAGCGGTAGTTTAAATCATACCTCAITATGA 1878
QY 1906 TCATTACCAATATTAATTTCTGTTGTTGATGATCACACATACAAAGCTCCAAATGG 1965
Db 1879 CCATTACCAATACATCAATTTGAGTGGTTGAGAGGCTTTNTGAGCAGCTAAGGG 1938
QY 1966 CTATACCTTGAAGATTTGTTGCGAGTAAAGTACTACGTAGAACACCTGACGAACG 2025
Db 1939 GTATACCTTGGAGTCTTTTGGCGACTGTCAAGTACTATGTGCAATCCAAACGAACG 1998
QY 2026 TCACATCTTAATGATGAGTGGGCAATGCCAGTACGATGTTGTTAGGCAAGAAACCA 2085
Db 1999 TCGCATTCAGATAATGGTTTGGTACGCTAGCGACCATGTTTCAAGAAACCAAAATGG 2058
QY 2086 CAGTGAAGATCCAAATAGAATCTTAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
Db 2059 TCAAGCTGATACCAATCAACGGAAGAAACCAAGCGAGGAGAAACCTCAGACAGAAACC 2118
QY 2146 TGAGCCAGAAAGTCCCTCAAGTAGAGA 2171
Db 2119 TGAGGAAGAAACCCCTCGAGAAGAGA 2144
```

RESULT 8

```
US-09-536-784-65
; Sequence 65, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
; US-09-536-784-65
```

```
Query Match 41.3%; Score 987.6; DB 4; Length 2290;
Best Local Similarity 67.7%; Pred. No. 1.2e-263;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA---TAATCGTGTTC 57
Db 4 TTCTTATGAACCTTGGTCGTCAACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC 63
QY 58 CTATATAGATGGAACAACAGCAGCGCAAAACGAGAAATTCACCTCCTGATGAGGTTAG 117
Db 64 TTATATAGATGGTGTGATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACCAAGATGAAGTCAG 123
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAACTCGTCATCAAGATAACAGACCAAGCGCTATGT 177
Db 124 TAAGAGGGAGGGATCAACGCCGACAAATNGTNAICAAAGATTACGATCAAGGTTATGT 183
QY 178 CACTTCATATGCGACCACTATCATTTATTAATGTTAAGGTTTCCTTATGACGCTATCAT 237
Db 184 GACCTCTCATGGAGACCAATTATCACTAATAATGCAAGGTTCCCTTATGATGCCATCAT 243
QY 238 CAGTGAAGAATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGGATATTGTTAA 297
Db 244 CAGTGAAGAGCTCCTCATGAAAGATCCGAAATTTATCAGTTTGAAGGATTCAGACATTTGTCAA 303
QY 298 TGAGGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGCA 357
Db 304 TGAATCAAGGGTGGTATGTCATTAAGGTAACCGTAATTAATCTATGTTTACCTTAAGCA 363
QY 358 TGCTGCCCAACGCGATAACGTCCTGATCAAAAGAGGAAATCAATCGACAAAAACAAGAGCA 417
Db 364 TGCAGCTCATGCGGATAATATTTCGGACAAAAGAGAGATTAAACGTCGAGAACGAGAACG 423
QY 418 TAGTCAACATCGTGAAGGTGGAACTCCCAAGAAACCATGCTGCTGTTGCCCTTGGCAGCTTC 477
Db 424 CAGTCATATCAT-----AACTCAAGAGCAGATAATGCTGTTGTCGACCCAGAGC 474
QY 478 GCAAGGACGCTATCTACAGATGATGTTTATCTTTTAACTCTTCTGATATCATGAGAGA 537
Db 475 CCAAGGACGTTATACAAAGGATGATGGGTATATCTTCAATGATCTGATATCATTTAGAGA 534
QY 538 TACTGCTGATGCTTATATCGTTCCCTCATGAGATCAATACCATTAATCCCTAAGAATGA 597
Db 535 CACGGGTGATGCTTATATCGTTTCTCTCAACGGGACCATTAACCATTAATCCCTAAGAATGA 594
QY 598 GTTATCAGCTAGGAGTTGGCTGCTGCGAAGCCCTTCTATCTGTCGAGGAATCTGTC 657
```

D	b	595	GTTTATCAGCTAGCGAGTTAGCTGCTGCAGAAGCCCT-----	629
Q	y	658	AAATTCAAGAACCTATTCGGCGACAAATAATACGATATACACTTCAAGAACAAAACCTGGGTACC	717
D	b	630	-----AATTGGAATTGGGAAGCAGGGATCTCGTCTCTTTCAAGTTCTTGATTATAATATGC	681
Q	y	718	TTCGTGAAGCAATCCAGGAACATAAATACTAACACAAGCAACAACAGCAACAACACTAACAG	777
D	b	682	AAATCCAGCTCAACCAAGATTGTGAGAGAACCAATCTGACTGTCACTCCAACITATCA	741
Q	y	778	TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTTGAAAAAGCTCTACAAAAGTGCCTTT	837
D	b	742	TCA---AAATCAAGGGGAAAAACATTTCAAGCCCTTTTACGTGAATTGTATGCTTAAACCCCTT	798
Q	y	838	GAGTCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAGTCG	897
D	b	799	ATCAGAAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAATCACAGTCG	858
Q	y	898	AACAGCTPAGAGGTGTTTCAGTGCACACGGAGATCATTAACCACTTCATCCCTTACTCTCA	957
D	b	859	AACCGCACAGGTGTAGCTGTCCCTCATGTATACCATTACCACCTTTATCCCATTATGAACA	918
Q	y	958	AATGCTTGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTCCTTCTGTTATCGTTCAAACA	1017
D	b	919	AATGCTCTGAATTTGAAAAACGAATTTGCTGTATTAATTCCTCCTTCTGTTATCGTTCAAACA	978
Q	y	1018	TTGGGTACAGATTCAGAGCCAGAACAAACAAGTCCACACCGACTCCGGMACCTAGTCC	1077
D	b	979	TTGGGTACCAAGATTCAGACCGAACCAACCAAGTCCCAATTCGACTCCGGAACCTAGTCC	1038
Q	y	1078	AGGCCCGCAACTCTCACCAATCTTAAAAATAGACTCAAAATCTTTC-----TTTT	1125
D	b	1039	AAGTCCGCACTGACCAAAATCTCAACAGCTCCAGCAATCCAATTCATGAGAAAT	1098
Q	y	1126	GGTTTAGTCAGCTGTACGAAAGTTGGGAAGGATATGTATTCGAGAAAGGGCATCTC	1185
D	b	1099	GGTCAAGAAGCTCTTTCGAAAGTAGCGATGGTATGTCTTTTCAGAGAGAAATCGAGTTTC	1158
Q	y	1186	TCGTTATGTCTTTCCGGAAGATTTACCATCTGAAACGTTTAAAAATCTTGAAAGCAAGTT	1245
D	b	1159	TCGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGGAGGATTTATAGCAAACT	1218
Q	y	1246	ATCAAAACAAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATTTGCTCCTCTCG	1305
D	b	1219	GGCCAGCAGGAAAGTTTATCTATAGCTTAGGAGCTAAGAAACTGACCTCCCATCTAG	1278
Q	y	1306	TGACCAAGAATTTTATGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTT	1365
D	b	1279	TGATCGAGAAATTTTACAAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT	1338
Q	y	1366	TGNAAAATAAGGTCGTAAATCTTGATTTCCAGACCTTAGACAAATATTAGAACGCTTGAA	1425
D	b	1339	TGATAANTAAGGTGCACAGTTGATTTTGAGGCTTTTGGATTAACCTTTTGAAACGACTCAA	1398
Q	y	1426	TGATCAATPCGACTAATAAAGAAAAATTTGGTAGATGATTAATTTGGCAATTCCTAGCACCAAT	1485
D	b	1399	GGAATCNCAAGTATNAAGTCAAGTAGTGTGANGATATCTTGCCCTCTTAGCTCCGAT	1458
Q	y	1486	TACCCATCCAGAGCGACTTGGGAAACCAAATCTTCAAATTTAGTATATCTGAAAGCAGAT	1545
D	b	1459	TCGTATCCAGAACGTTTAGGAAAACCAAATTCGCAAAATTAACCTTACACTGATGATGAGAT	1518
Q	y	1546	TCGTATTTGCTCAATTTAGCTGATATAGTATACACGTTCAGATGGTTTACATTTTGTATGAACA	1605
D	b	1519	TCAAGTAGCCAAGTTGGCAGGCAAGTACACACAGACAGACGGTTATATCTTTGATCTCTCG	1578
Q	y	1606	TGATATAATCAGTGATGAAGAGATGCAATATGTAACGCTCATATGGGCCATAGTCACTG	1665
D	b	1579	TGATATAACCAAGTATGAGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTG	1638
Q	y	1666	GATTCGAAAGATAGCCCTTTCTGTATAAGGAAAAAGTTTCAGCTCAAGCTTACTATAAGA	1725

RESULT 9

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-527-243

Query Match 27.4%; Score 653.8; DB 4; Length 2359;
 Best Local Similarity 68.3%; Pred. No. 4.3e-171;
 Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

```

QY 1 TTCTTACGAGTTCGGGACTGTATCAAGCTAGAACGGTTTAAGGAAA-----TAATCGTGTTC 57
Db 939 TTCTTACGAGTTCGGGACTGTATCAAGCTAGAACGGTTTAAGGAAA-----TAATCGTGTTC 998
QY 58 CTATATAGATGGAAGAACCAAGCGACGCAAAACCGAGAAATTCGACTCCTCATGAGGTTAG 117
Db 999 TTATATAGATGTTGATCAGGCTGGTCAAAAGGCGAGAAATTCGACACGATGAGTCAG 1058
QY 118 CAAAGCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGT 177
Db 1059 TAAGAGGGAGGGGATCAACCGCCGAAACAAATGTTATCAAGATTACGATCAAGGTTATGT 1118
QY 178 CACTTCACATGGCGACCACTATCATTTATTAATGTTAGGTTCCCTATGACGCTATCAT 237
Db 1119 GACCTCTCATGGAGACCATTTATCAATTAATGGAAGGTTCCCTATGATGCCATCAT 1178
QY 238 CAGTGAAGAAATTAATCAATGAAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAA 297
Db 1179 CAGTGAAGACTCTCATGAAGATCGGAATATCATGTTGAAGATTTCAGACATTTGTCAA 1238
QY 298 TGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGA 357
Db 1239 TGAATCAAGGGTGGTATGTTATCAAGGTAAACCGTAAATCTATGTTTACCTTAAGGA 1298
QY 358 TGCTGCCACGCGGATAACGTCGTACAAAGAGGAAATCAATCGACAAACCAAGCA 417
Db 1299 TGAGCTCATGCGGATAATATTCGGAACAAAGAGAGATTAACGTCGAGAGCAGGAACG 1358
QY 418 TAGTCAACATCGTGAAGGTGGAACTCCAAGAACGATGGTGTGTTGCCCTTGCGACGTTTC 477
Db 1359 CAGTCATATCAATCT-----CAAGACGAGATATGCTGTTGCTGCGACCGAGGC 1409
QY 478 GCAAGACGCTATCTACAGATGATGGTTATATCTTTAATGTTCTTGATATCATAGAGA 537
Db 1410 CCAAGGACGTTATACAGCGATGATGGTATATCTTCAATGATCATCTGATATCATAGGA 1469
QY 538 TACTGTGTATGTTATATCTGTTCTCATGAGATGATTAACATTACATTCCTTAAGATGA 597
Db 1470 CACGGTGTATGTTATATCTGTTCTCATGAGGACCATTAACATTACATTCCTTAAGATGA 1529
QY 598 GTTATCAGCTAGCGAGTGTGGCTGCTGAGAGGCTTCTCTATCTGTTGCGAGGAAATCTGTC 657
Db 1530 GTTATCAGCTAGCGAGTGTAGCTGCTGAGAGGCTTATGGAATGG----- 1574
QY 658 AAATTCAGAACCTATCGCGGCAAAATAGCGATAACATTCAGAGAACAACTGGGTACC 717
Db 1575 -GAAGCAGGGATCTCGTCTCTTCTTCAAGTCTTAGTTTATAATGCAATCCAGCTCAACCA 1633
QY 718 TTCTGTGAACCAATCCAGGATCTACAACTAACACCAACCAACAGCAACACTAACAG 777
Db 1634 GATTGTGAGAGAACCAATCTGACTGT-----CACTCCAACTTA 1673
QY 778 TCAAGCAAGTCAAGATTAATGACATTTGATAGTCTCTTGAACAGCTCTCAAACTGCTTT 837
Db 1674 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTTGCTTAACCTTT 1733
QY 838 GAGTCAACAGATGTAGATCTGATGGCCTTGTCTTTGATCCAGCAAAATCACAAGTCG 897
Db 1734 ATCAGAACCCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAAAATCACAAGTCG 1793
QY 898 AACAGCTAGAGTGTTCGATGCCACACGAGATCATTTACCACTTCATCCCTTACTCTCA 957
Db 1794 AACCCGACAGGTTGAGCTGTCCTCATGGTAACCACTTATCCCTTATGAACA 1853
QY 958 AATGTCGAATTTGGAAGAACGAATCGCTCGTATTATTCCTCCCTTGGTTATCGTTCAACCA 1017

```

RESULT 10

```

US-08-961-083-181
; Sequence 181, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-181

Query Match 16.1%; Score 385.4; DB 3; Length 1342;
Best Local Similarity 65.3%; Pred. No. 8.1e-97;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

Qy 37 TAAGGAAAATAATCGTGTTCCTATATAGATGGAAGAAACGAGCGAGAAACCGAGAA 96
Db |||||
Qy 28 TAAGGACAAATAATCGTGTCTTATGTGATGGAGCGAGTCAAGTCAAGAAAGTGAA 87
Db |||||
Qy 97 TTGACCTCTGATCAGGTAGCAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 156
Db |||||
Qy 88 CTTGACACCAAGCAGGTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATTTGTAACAA 147
Db |||||
Qy 157 GATAACAGCAAGGCTATGCTTACATGAGGTAAGGAATCAATGCTGAGCAAAATCGTCATCA 216
Db |||||
Qy 148 AATTACAGATCAGGCTATGTAACGTCACCGGTGACCACTATCATTAATAATGGGAA 207
Db |||||
Qy 217 GGTTCCTTATGACGCTATGTAACGTCACCGGTGACCACTATCATTAATAATGGTAA 276
Db |||||
Qy 217 GGTTCCTTATGACGCTATGTAACGTCACCGGTGACCACTATCATTAATAATGGTAA 276
Db |||||
Qy 208 AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGATCCAACTATCACT 267
Db |||||
Qy 277 AAAGATCAGGATATTTGTAATGAGGTCAAGGCTGAGATGTTATCAAGGTAGTGGAA 336
Db |||||
Qy 268 TAAAGACGCTGATATTTGTCATGAAGTCAAGGCTGTTATATCATCAAGTTCGATGAA 327
Db |||||
Qy 337 ATACTATGTTTACCTTAAGGATGCTGCCACCGGATAACGTCCTGACAAAGAGAAAT 396
Db |||||
Qy 328 ATATATGCTACCTGAAGATGACGCTCATGCTGATAATGTTGCAACTAAAGATGAA 387
Db |||||
Qy 397 CAATCGCAAAACAAAGAGATGATCAACATGCTGAAGTGAAGTGAAGTGAAGTGAAG 456
Db |||||
Qy 388 CAATCGTCAAAACAAAGAACATGTCAAGATAATGAGAAGTTA-----ACTC 435
Db |||||
Qy 457 TGCTGTGCTTGGCCTTGGCAGCTTGGCAGGAGCTTACTACAGATGATGTTATCTTTAA 516
Db |||||
Qy 436 TAATGTTCTGTGGAAGGCTCTCAGGAGCATATACGCAAAATGATGTTATGTTTAA 495
Db |||||
Qy 517 TGCTTCTGATATCATAGAGTACTGTGTGCTTATATCGTTCCTCATGAGATCAATTA 576
Db |||||
Qy 496 TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGCTCTCATGAGGTCATTA 555
Db |||||
Qy 577 CCAATTAATTCCTAAGATGATGATCAGTACGAGTGGCTGCTGAGAGCTTCTCT 636
Db |||||
Qy 556 TCACTACATTCCTCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGTAAAGCATCT 615
Db |||||
Qy 637 ATCTGGTCGAGAAATCTGTCAAATTCAGAACCTATCGCCGACAAAATAGCGATAACAC 696
Db |||||
Qy 616 GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGCTAGTGACA 672
Db |||||
Qy 697 TTCAAGAACAACTGGGTACCTTCTGTGAAGCAATCCAGAACTCAAAATCTAACAACAG 756
Db |||||
Qy 673 -----TAAACGCAATCTGT 687
Db |||||
Qy 757 CAACAGACACACATCAAGTCAAGCAAGTCAAGCAATGATGATGATGATGATGATGAT 816
Db |||||
Qy 688 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAAATCTGAAAATCTCCAGAGCTTTTGA 747
Db |||||
Qy 817 ACAGCTCTACAACTGCCTTTTGGTCAACGACATGTAGAATCTGATGCTGCTGCTTGA 876
Db |||||
Qy 748 GGAATCTTATGATTCACCTAGGCCCAAGCTTACAGTGAATCAGATGGCTGCTTGA 807
Db |||||
Qy 877 TCCAGCAAAATCAAGTTCGAAACAGCTAGAGGTGTCAGTGGCCACACGAGATCAATTA 936
Db |||||
Qy 808 CCGCTCTAAGATTTATCAGTCGTACACCAAAATGGAGTTGGATTCGCGATGGCGACATTA 867
Db |||||
Qy 937 CCATTCATCCCTTACTCTCAATGCTGAATTTGGAAGAACGAATCGCTCGTATTATCC 996
Db |||||
Qy 868 CCATTTTATCTCTACAGCAAGCTTTCTGCCTTAGAAGAAAGATTCAGAGATGGTGCC 927
Db |||||

Qy 997 CCTTCGT 1003
Db 928 TATCAGT 934

RESULT 11

US-09-536-784-181
; Sequence 181, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 181:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-536-784-181

Query Match 16.1%; Score 385.4; DB 4; Length 1342;
Best Local Similarity 65.3%; Pred. No. 8.1e-97;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

Qy 37 TAAGGAAAATAATCGTGTTCCTATATAGATGGAAGAAACGAGCGAGAAACCGAGAA 96
Db |||||
Qy 28 TAAGGACAAATAATCGTGTCTTATGTGATGGAGCGAGTCAAGTCAAGAAAGTGAA 87
Db |||||
Qy 97 TTGACCTCTGATCAGGTAGCAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 156
Db |||||
Qy 88 CTTGACACCAAGCAGGTAGCCAGAAAGAGGAATTCAGGCTGAGCAAAATTTGTAACAA 147
Db |||||
Qy 157 GATAACAGCAAGGCTATGCTTACATGAGGTAAGGAATCAATGCTGAGCAAAATCGTCATCA 216
Db |||||
Qy 148 AATTACAGATCAGGCTATGTAACGTCACCGGTGACCACTATCATTAATAATGGGAA 207
Db |||||
Qy 217 GGTTCCTTATGACGCTATGTAACGTCACCGGTGACCACTATCATTAATAATGGTAA 276
Db |||||
Qy 208 AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGATCCAACTATCACT 267
Db |||||
Qy 277 AAAGATCAGGATATTTGTAATGAGGTCAAGGCTGAGATGTTATCAAGGTAGTGGAA 336
Db |||||
Qy 268 TAAAGACGCTGATATTTGTCATGAAGTCAAGGCTGTTATATCATCAAGTTCGATGAA 327
Db |||||
Qy 337 ATACTATGTTTACCTTAAGGATGCTGCCACCGGATAACGTCCTGACAAAGAGAAAT 396
Db |||||

Db 328 ATATTATGCTCTACCTGAAAGATGCGAGCTCATGCTGATATGCTGAACTTAAGATGAAT 387
 QY 397 CAATCGACAAAACAAAGAGATAGTCAACATCGTGAAGGTGAACTTCCAAAGAAACGATGG 456
 Db 388 CAATCGTCAAAAACAAAGACATGTCAAAGATATGAGAGGTTA-----ACTC 435
 QY 457 TGCTGTGCTTGGCAGCTTCCAGGAGCGCTATATCTACAGATGATGGTTATATCTTTAA 516
 Db 436 TAATGTGCTGTAGCAAGGTCTCAGGAGCGATATACGACAAATGATGGTTATGCTTTAA 495
 QY 517 TGCTTCTGATATCATAGAGGATCTGCTGATGCTTATATCGTTCTCATGGAGATCATTA 576
 Db 496 TCCAGCTGATATATCGAAGATACGGGTAAATGCTTATATCGTTCTCATGGAGTCACAT 555
 QY 577 CAATTACATCTTAAAGATGATGATATACGATGAGGTGGCTGCTGCGAAGCCCTTCT 636
 Db 556 TCACGTACATTTCCAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 615
 QY 637 ATCTGCTCGAGGAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATACAC 696
 Db 616 GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTCAACAGCTTAGTGCAA 672
 QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATACAAATATCTAACACA 756
 Db 673 -----TAACAGCAATCTGT 687
 QY 757 CAACACAGCAACACTATACAGTCAAGCAAGTCAAAAGTAAATGATGATGATCTTTGAA 816
 Db 688 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAATCTGAAATCTCCAGAGTCTTTTGA 747
 QY 817 ACAGCTCTACAAACTGCTTTGAGTCAAGCATGATGAGATCTGATGCGCTTGTCTTTGA 876
 Db 748 GGAATCTATGATTCACCTAGCGCCCAAGTTACAGTGAATCAGATGCGCTGCTTTGA 807
 QY 877 TCCAGCAAAATCAAGTCAAGTCAAGTCAAGTGTGTCAGTGGCCACACGAGATCATTA 936
 Db 808 CCTGCTAAGATTAATCAGTTCGTACACCAATGAGTTGCGATTCCGATGGGACCATTA 867
 QY 937 CCATTCATCCCTTACTCTCAAAATCTGAAATGGAAGCAAGATGCGTTCGATTTATTC 996
 Db 868 CCACCTTATTCCTTACAGCAAGCTTTCTGCTTTAGAGAAAAGATTGCCAGAAATGGTGCC 927
 QY 997 CCTTCGT 1003
 Db 928 TATCAGT 934

RESULT 12
 US-09-468-656A-7
 ; Sequence 7, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John B.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-7

Query Match 16.1%; Score 385.4; DB 4; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 8.4e-97;

Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAGAAATATATCGTGTTCCTTATATAGATGAAACAAAGCGACGCAAAAAACGAGAA 96
 Db 99 TAAGACAATATATCGTGTCTCTTATGTGTGATGGCAGCCAGTCAAGTGCAGAAAGTGA 158
 QY 97 TTTGACTCTGATGAGGTTAGCAAGCGTGAAGGAATCAATGCTGAGCAATCGTCATCAA 156
 Db 159 CTTGACACAGACACAGGTTAGCCGAAAGAGGAATTCAGGCTGAGCAATTTGTAATCAA 218
 QY 157 GATAACAGACCAAGGCTATGTCACTTTCATGCGCACCACTATCATTTATTAACAATG 216
 Db 219 AATTAACAGATCAGGCTATGTAACGTCAACGTCACACATCATCTACTTATTAATGG 278
 QY 217 GGTTCCTTATGACGCTATCATCAGTGAAGAAATTAATCTATGAAGATCCAAATATAAG 276
 Db 279 AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTTGATGAAGGATCCAAATCATCA 338
 QY 277 AAAAGATGAGGATATTTGTAATGAGGTCAAGGTCGATATGTTATCAAGGTAGATGAAA 336
 Db 339 TAAAGACGCTGATATTTGTAATGAAGTCAAGGTCGTTATATCATCAAGGTCGATGAAA 398
 QY 337 ATACTATGTTTACCTTAAAGGATGCTGCCACGCGATTAACGTCCTTACAAAAGAGAA 396
 Db 339 ATATTATGCTTACCTGAAAGATGACGCTCATGCTGATTAATGTTTGAATTAAGATGA 458
 QY 337 CAATCGACAAAACAAAGAGATAGTCAACATGCTGTAAGGTGGAATCCCAAGAAACGAT 456
 Db 459 CAATCGTCAAAAACAAAGAACATGTCAAAGATAATGAGAAGGTTA-----ACTC 506
 QY 457 TGCTGTGCTTGGCAGCTTCCAGAGGACGCTATATCTACAGATGATGTTATATCTTTAA 516
 Db 507 TAATGTTGCTGTAGCAAGGTTCTCAGGAGCATATACGACAAATGATGTTATGTTTAA 566
 QY 517 TGCTTCTGATATCATAGAGGATCTGCTGATGCTTATATCGTTCTCATGAGATCATTA 576
 Db 567 TCCAGCTGATATATCGAAGATACGGTAAATGCTTATATGTTTCTCATGGAGTCACTA 626
 QY 577 CCATTACATCTTAAAGATGATTAATCAGTGAAGGTTGCTGCTGAGAAAGCCCTTCT 636
 Db 627 TCACGTACATTTCCAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCACAT 686
 QY 637 ATCTGCTCGAGGAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATACAC 696
 Db 687 GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGCTAGTGCAA 743
 QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATATCTAACACA 756
 Db 744 -----TAACAGCAATCTGT 758
 QY 757 CAACACAGCAACACTATACAGTCAAGCAAGTCAAAAGTAAATGATGATGATCTTTGAA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAATCTGAAATCTCCAGAGTCTTTTGA 818
 QY 817 ACAGCTCTACAACTGCTTTGAGTCAAGCATGATGAAATCTGATGCGCTTGTCTTTGA 876
 Db 819 GGAATCTTATGATTCACCTAGGCCCCAACGTTACAGTAATCAGATGCGCTGCTTTGA 878
 QY 877 TCCAGCAAAATCAAGTCCGAAACAGCTAGAGGTGTTGAGTGGCCACACGAGATCATTA 936
 Db 879 CCCTGCTAAGATTAATCAGTCTGTAACCAATAGGAGTTGCGATTCCGATGGGACCATTA 938
 QY 937 CCATTCATCCCTTACTCTCAAATGCTGAAATGGAAGAAAGATTCGCTTCGATTTATTC 996
 Db 939 CCACCTTATTCCTTACAGCAAGCTTTCTGCTTTAGAGAAAAGATTGCCAGAAATGGTGCC 998
 QY 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 13
 US-08-961-527-192/c

```

; Sequence 192, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-192

```

```

Query Match 16.1%; Score 385.4; DB 4; Length 6867;
Best Local Similarity 65.3%; Pred. No. 2e-96;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATTAATCGTCTTCTATATAGTGGAAACAAAGCAGCGCAAAACCGAGAA 96
DB 6687 TAAGGACATTAATCGTCTTCTATATAGTGGAGGAGGAGGAGGAGGAGGAGG 6628

QY 97 TTGACTCTCTGATGAGGTAGCAAGCGGTGAAGGAAATCAATGCTGAGCAATCGTCATCA 156
DB 6627 CTTGACACCAAGCAGGTAGCCAGAAAGAGGAAATTCAGGCTGAGCAATTTGTAATCAA 6568

QY 157 GATAACAGACCAAGGCTATGTCATTCACATGGCGGACCACTATCAATTAACATGCTAA 216
DB 6567 AATTACAGATCAGGCTATGTAAGCTCAGCGGTGACCACTATCAATTAACATGCTAA 6508

QY 217 GGTTCCTTACGCTATCACTGATGAGGAAATTAATCAAGGAGGAGGAGGAGGAGG 276
DB 6507 AGTTCCTTATGATGCTCTTCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6448

QY 277 AAAAGATGAGGATTAATGATAGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 336
DB 6447 TAAAGACGCTGATATGTCATGAAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 6388

QY 337 ATACTATGTTTACCTTAAGGATGCTCCCGCGGATTAAGTCCGTCAGGAGGAGGAGG 396
DB 6387 ATATATGTTTACCTTAAGGATGCTCCCGCGGATTAAGTCCGTCAGGAGGAGGAGG 6328

QY 397 CAATCGCAAAACAAAGAGGATGATCAACATCGTGAAGGAGGAGGAGGAGGAGGAGG 456
DB 6327 CAATCGTCAAAACAAAGAGGATGATCAACATCGTGAAGGAGGAGGAGGAGGAGGAGG 6280
; ACTC
QY 457 TGCTGTTGCCCTTGGCAGCTTCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 516
; TATACTTAA

```

RESULT 14

```

US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504

```

```

DB 6279 TAAATGTTGCTAGCAAGGTCTCAGGAGCGATATACGACAAATGATGTTATGCTTTAA 6220
QY 517 TCGTTCTGATATCATAGAGGATCTGGTGTATATGCTTATATCGTTCCATGAGGATCATTA 576
DB 6219 TCCAGCTGATATTAATCAAGATACGGGTAAATGCTTATATCGTTCCATGAGGATCATTA 6160
QY 577 CCATTACATTCCTAAGAATGAGTTATCAGCTAGCAGGTTGGTGTGTCGAGAACCTTCCT 636
DB 6159 TCATACATTCCTCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCAGATCT 6100
QY 637 ATCTGTCGAGGAAATCTGTCAAAATCAAGAACCTATCGCCGACAAATAGCGATACAC 636
DB 6099 GGCTGAAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTAGCAAA--- 6043
QY 697 TTCAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATACAAATACTAACACAAG 756
DB 6042 -----TACACGCAATCTGT 6028
QY 757 CAACAACAGCAACACTAACAAGTCAAGCAAGTCAAAAGTAAATGACATTTGATGTTGAA 816
DB 6027 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATAATCTGAAATCTCCAGAGTCTTTTGA 5968
QY 817 ACAGCTCTCAAAACTGCTTTGAGTCAACGACATGATAGTATCTGATGGCCTTCTTTGA 876
DB 5967 GGAATCTATGATTACCTAGCGCCCAACGTTACAGTGAATCAGATGGCCTGCTTTGA 5908
QY 877 TCCAGCACAAATCACAAGTCGAACAGCTAGAGGTGTTGCAAGTCCACACGAGATCATTA 936
DB 5907 CCTGCTAGATTAATCAGTGTACCAAAATGAGGTTGCGATTCGCGATCGCAGCATTA 5848
QY 937 CCATCTTATCCTTACTCTCAAAATGCTGAATTTGGAAGAACGAATCGTCTGATTTATTC 996
DB 5847 CCATTTTATCTTACAGCAAGCTTTCTGCTTTAGAAGAAAGATGCCAGAAATGGTGCC 5788
QY 997 CTTGCT 1003
DB 5787 TATCAGT 5781

```


Db 600 CAGAGGAATCAGAAAGACCTCAGGTCGAGACTGAAAGGTTGAAGAAACTGAGAGAG 659
 QY 2202 CAGAAGTTTGGTGGAAAGTAACGATCTAGTCTGAAGCCATGCAACAGAACTC 2261
 Db 660 CTGAAGATTACTTGGAAAAATCCAGATCCAAATTATCAAGTCCAATGCGCAAGAGACTC 719
 QY 2262 TAGCTGTTTACGAAATAATTTGACTCTTCCAAATTATGGATAACAATAGTATCATGCGAG 2321
 Db 720 TCACAGATTAAAAATAATTTACTATTGCGACCCAGGACCAATACTATTATGGCAG 779
 QY 2322 AAGCAGAAAAATTAATCTGCTTGTAAAGGAAAGTAA 2358
 Db 780 AAGCTGAAACACTATTGCTTTATTAAGGAGAGTAA 816

RESULT 16
 US-08-743-637B-34/c
 ; Sequence 34, Application US/08743637B
 ; Patent No. 5994066
 ; GENERAL INFORMATION:
 ; APPLICANT: BERGERON, Michel G.
 ; APPLICANT: PICARD, Francois J.
 ; APPLICANT: OUELLETTE, Marc
 ; APPLICANT: ROY, Paul H.
 ; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
 ; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
 ; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
 ; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
 ; NUMBER OF SEQUENCES: 273
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: QUARLES & BRADY
 ; STREET: 411 EAST WISCONSIN AVENUE
 ; CITY: MILWAUKEE
 ; STATE: WISCONSIN
 ; COUNTRY: USA
 ; ZIP: 53202-4497
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/743,637B
 ; FILING DATE: 04-NOV-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/526,840
 ; FILING DATE: 11-SEP-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BAKER, Jean C.
 ; REGISTRATION NUMBER: 35,433
 ; REFERENCE/DOCKET NUMBER: 850586.90012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (414) 277-5000
 ; TELEFAX: (414) 277-5591
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 841 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; US-08-743-637B-34

Query Match 4.6%; Score 109; DB 2; Length 841;
 Best Local Similarity 65.4%; Pred. No. 2.7e-20;
 Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;
 QY 2090 GAAGATCCAAATAAGAACTTCAAGCGGATGAGAGCCAGTAGAGGAACACCTGCTGAG 2149
 Db 633 GAGTCTCCAAAACCAACAGAGGAACCAAGAAAGAAATCACCAGAAATCACCAGAGAA 574

QY 2150 CCAGAAAGTCCCTCAAGTAGAGACTGAAAGAGTAGAACCCCAACTCAAGAGCAGAGATT 2209
 Db 573 TCAGAGAAAGACCTCAGGTCGAGACTGAAAGGTTAAGAA--AACTGAGAGAGGCTGAAGAT 517
 QY 2210 TTGCTTGGAAAGTAACGATCTAGTCTGAAGCCAAATGCAACAGAACTCTAGCTGGT 2269
 Db 516 TTACTTTGAAAAATCCAGAAATCCAAATTATCAAGTCCAAATGCCAAAGAGACTCTCACAGGA 457
 QY 2270 TTACGAAATAATTTGACTCTTCAAAATTTATGATAACAATAGTATCATGCGAGAGCAGAA 2329
 Db 456 TTAATAAAATAATTTACTATTGCGACCCAGGACCAATACTATTATGGCAGAGCTGAA 397
 QY 2330 AAATTACTTGCCTTGTAAAGGAAAGTAA 2358
 Db 396 AAATCTATGCTTTATTAAGGAGAGTAA 368

RESULT 17
 US-08-526-840B-34/c
 ; Sequence 34, Application US/08526840B
 ; Patent No. 6001564
 ; GENERAL INFORMATION:
 ; APPLICANT: BERGERON, Michel G.
 ; APPLICANT: OUELLETTE, Marc
 ; APPLICANT: ROY, Paul H.
 ; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
 ; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
 ; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
 ; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
 ; NUMBER OF SEQUENCES: 177
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: QUARLES & BRADY
 ; STREET: 411 East Wisconsin Avenue
 ; CITY: Milwaukee
 ; STATE: Wisconsin
 ; COUNTRY: USA
 ; ZIP: 53202-4497
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/526,840B
 ; FILING DATE: 11-SEP-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/304,732
 ; FILING DATE: 12-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BAKER, Jean C.
 ; REGISTRATION NUMBER: 35,433
 ; REFERENCE/DOCKET NUMBER: 850586.90012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (414) 277-5000
 ; TELEFAX: (414) 277-5591
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 841 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; US-08-526-840B-34

Query Match 4.6%; Score 109; DB 3; Length 841;
 Best Local Similarity 65.4%; Pred. No. 2.7e-20;
 Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;
 QY 2090 GAAGATCCAAATAAGAACTTCAAGCGGATGAGAGCCAGTAGAGGAACACCTGCTGAG 2149

Db 633 GAGTCTCCAAAACCAACAGAGGACCAAGAGAAATCACCAGAGAAATCACCAGAGAA 574
 Qy 2150 CCAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAGAGAGAGTT 2209
 Db 573 TCAGAGAACTCAGGTCGAGACTGAAAGGTTAAGAA--AACTGAGAGAGGCTGAAGAT 517
 Qy 2210 TTGCTTGCAGAAAGTAACGAGTTCTAGTCTCAAGCCCAATCAGAGAGAACTCTAGCTGGT 2269
 Db 516 TTACTTGGAAAAATCCAGATCCAAATATCAAGTCCAAATGCCAAAGAGACTCTCAGAG 457
 Qy 2270 TTACGAAATTAATTTGACTCTTCAAAATATGGAATACAAATAGTAGTGGCAGAGAGAGAA 2329
 Db 456 TTAAGAAATTAATTTACTATTGTCAGCCAGCAACAATACTATTATGGCAGAGAGCTGAA 397
 Qy 2330 AAATTACTTGGTGTGTAAGAGAGTAA 2358
 Db 396 AAATATTGGCTTTATTAAAGAGAGAGTAA 368

RESULT 18

US-08-232-463-14/c
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367
 ; GENERAL INFORMATION:
 ; APPLICANT: DORNER, F.
 ; APPLICANT: SCHEIFLINGER, F.
 ; APPLICANT: FALKNER, F. G.
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 ; NUMBER OF SEQUENCES: 52
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22313-0299
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/232,463
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/935,313
 ; FILING DATE:
 ; APPLICATION NUMBER: EP 91 114 300.6
 ; FILING DATE: 26-AUG-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703)836-9300
 ; TELEFAX: (703)683-4109
 ; TELEX: 899149
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7218 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; CLONE: pTZgpt-Fls
 ; US-08-232-463-14

Query Match 2.4%; Score 57.4; DB 1; Length 7218;
 Best Local Similarity 3.4%; Pred. No. 1.7e-05;
 Matches 13; Conservative 224; Mismatches 150; Indels 0; Gaps 0;

Qy 1974 TGGAAGATTGTTTGGCAGCAATTAAGTACTAGTAGAAACACCTGACGAAACGTCACATT 2033
 Db 1450 TAGAAGAAATTTGTACRR 1391
 Qy 2034 CTAATGATGCGGCGCAATGCCAGTCAGCATGTGTAGGCAAGAAAGACACAGTGAAG 2093
 Db 1390 RRR 1331
 Qy 2094 ATCAAATAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGAGAAACACCTGCTGAGCCAG 2153
 Db 1330 RRR 1271
 Qy 2154 AAGTCCTCAAGTAGAGACTGAAAAAGTAGAGCCCAACTCAAGAGAGAGAGTTTTCG 2213
 Db 1270 RRR 1211
 Qy 2214 TTGCGAAAGTAACGAGTTCTAGTCTCAAGCCCAATGCAACAGAAACTCTAGCTGGTTTAC 2273
 Db 1210 RRR 1151
 Qy 2274 GAAATAATTGACTCTTCAAATATGATAACAATATGATATCATGGCAGAGAGAGAGAAAT 2333
 Db 1150 RRR 1091
 Qy 2334 TACTTGGTGTGTTAAAGGAAGTAATC 2360
 Db 1090 RRR 1064

RESULT 19
 US-09-134-001C-2243
 ; Sequence 2243, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 2243
 ; LENGTH: 11091
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-2243

Query Match 2.2%; Score 51.6; DB 4; Length 11091;
 Best Local Similarity 47.3%; Pred. No. 0.00088;
 Matches 187; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

Qy 1096 AAATCTTAAATAGACTCAAATCTCTTTGGTTAGTCAGTCGAGTACGAAAGTTGGGA 1155
 Db 8481 AAAAGCCAAAAATGATGTAAATCAATCTCAAACTAATCAGCAAGTTGAAATGTCGAGCA 8540
 Qy 1156 AGGATATGTTATTCGAAGAAAAAGGGCATCTCTGTTATCTCTTTCGAAA--GATTACC 1212
 Db 8541 AAATAGTTTAGTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 8600
 Qy 1213 ATCTGAAACTGTAAAAATCTTGAAGCAAGTTATCAAAACAGAGAGTGTTCACACAC 1272
 Db 8601 AGCTGAAATCTTAAAGCGCAACAAAACAAATTCATGAAATAGAGCAAGATTTAGTGC 8660
 Qy 1273 TTAACTGCTAAAAAGAAATGTTGCTCTCGTCGACCAAGATTTTATGATAAGACATA 1332
 Db 8661 TACAAGAGAGAAAAAGCAATGCTTTACAAACATTTAGATGAACAGGTTAAAGAAATCAT 8720
 Qy 1333 TAATCTGTTAAGGCTCATAGGCTGTTGTTGNAATAAGGCTGTAATTTCTGATT 1392

Db 8721 TAATTTCTATAAATCAACGTAATACAGATAATGAGTAGATAATGCTAAACATCTTCGGTT 8780
QY 1393 CCAAGCCTTAGACAAATATTAGACGCTTGGAATGATGAATCGACTAATTAAGAAAAAT 1452
Db 8781 GAATAACATAACTGAATACAGACAGACAGATAATATAAAGAAAAATGCTATATTTAAAT 8840
QY 1453 GGTAGATGATTATTGGCATTCCTTAGCACCAATTA 1487
Db 8841 ATATGATGTTTCAGATACTCAAGAAAGCTATAATTA 8875

RESULT 20
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-2

Query Match 1.9%; Score 45.2; DB 1; Length 2277;
Best Local Similarity 29.7%; Pred. No. 0.022;
Matches 77; Conservative 51; Mismatches 131; Indels 0; Gaps 0;

QY 279 AAGATGAGGATATTGTTAATGAGTCAAGGTGGATATGTTATCAAGGTAGATGAAAT 338
Db 671 ARGARGAGATGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGARGARG 730
QY 339 ACTATGTTTACCTTAAGGATGCTGCCACGGGATAACGTCCTACAAAAGAGGAAATCA 398
Db 731 AYGGNGTNTTYGAYGARGARGARGARAAYATHGARWSNARGTNAACNARC 790
QY 399 ATCGACAAAACAAGACGATAGTCAACATCGTGAAGGTGGAACTCCAAAGAACGATGTG 458
Db 791 CNGTNCARATHCARAARMGNCNGTNAARMGNCNGCNCNCCNARWSNNGAYCAYW 850
QY 459 CTGTTGCCCTTGGCAGCTTCGCAAGACGCTATACAGATGATGTTTATCTTTAATG 518
Db 851 SNGARGARGAYWSNGAYTNGARGARWSNATHGAYGAYGNGARGARYTNGCNC 910
QY 519 CTTCGTGATATCATAGGA 537

US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:

Db 911 ARWSNGAYACNWSNACNGA 929

RESULT 21
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-2

Query Match 1.9%; Score 45.2; DB 1; Length 2277;
Best Local Similarity 29.7%; Pred. No. 0.022;
Matches 77; Conservative 51; Mismatches 131; Indels 0; Gaps 0;

QY 279 AAGATGAGGATATTGTTAATGAGTCAAGGTGGATATGTTATCAAGGTAGATGAAAT 338
Db 671 ARGARGAGATGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGARGARG 730
QY 339 ACTATGTTTACCTTAAGGATGCTGCCACGGGATAACGTCCTACAAAAGAGGAAATCA 398
Db 731 AYGGNGTNTTYGAYGARGARGARGARAAYATHGARWSNARGTNAACNARC 790
QY 399 ATCGACAAAACAAGACGATAGTCAACATCGTGAAGGTGGAACTCCAAAGAACGATGTG 458
Db 791 CNGTNCARATHCARAARMGNCNGTNAARMGNCNGCNCNCCNARWSNNGAYCAYW 850
QY 459 CTGTTGCCCTTGGCAGCTTCGCAAGACGCTATACAGATGATGTTTATCTTTAATG 518
Db 851 SNGARGARGAYWSNGAYTNGARGARWSNATHGAYGAYGNGARGARYTNGCNC 910
QY 519 CTTCGTGATATCATAGGA 537
Db 911 ARWSNGAYACNWSNACNGA 929

RESULT 22
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:

QY 1755 CAGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTACAACTGGTGAAG 1814
 Db 425 AAGATTTAAACACTAAACAACTATAAGTAATCAAGAGCGTTACACCTGATTTGCAAG 484
 QY 1815 GGGAAAAACGAATCCCACTCGTTCGAGCTTCCATATATGGTTGAGCATACAGTTGAGGTTA 1874
 Db 485 AGATAAATCAGTGGTAAATGTTCAACCACTAATGAGGAAAAACAAAAAGGTAGATGCCA 544
 QY 1875 AAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATAATATTAAATTTCTTGGT 1934
 Db 545 AAACGTAATCAACATACATTAATGTTAAAGTGTATCAAGAGTAATGATGAAAACTC 604
 QY 1935 TTGATGATCACATACAGCACTCAATGCGCTATA 1970
 Db 605 TTGTTGATAACAATAGTAATTCAAATAAGAAAAATA 640

RESULT 25

US-08-956-171E-454/C
 ; Sequence 454, Application US/08956171E
 ; Patent No. 6593114
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; Gil H. Choi
 ; Patrick S. Dillon
 ; Craig A. Rosen
 ; Steven C. Barash
 ; Michael R. Fannon
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 5256
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/956,171E
 FILING DATE: 20-Oct-1997
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861
 FILING DATE: January 5, 1996
 APPLICATION NUMBER: 08/781,986
 FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 454:

SEQUENCE CHARACTERISTICS:
 LENGTH: 4358 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 454:

US-08-956-171E-454

Query Match 1.8%; Score 42.4; DB 4; Length 4358;
 Best Local Similarity 47.1%; Pred. No. 0.19;
 Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
 QY 1695 AAAAAGTTGAGCTCAAGCTATATCTAAAGAAAAAGGTATCTCTACCTCCATCTCCAGAG 1754

Db 599 ATAAAGTTGAAGTTTCAACTGCCAAATCAGATGAGCAAGCTTACCCTAAATCTACGAATG 540
 QY 1755 CAGATGTTAAAGCAAAATCCCAACTGGAGATAGTCAGCAGCTATTACAAATCGTGTGAAAG 1814
 Db 539 AAGATTTAAACACTAAACAACTATAAGTAATCAAGAAAGCGTTACAACTGATTTGCAAG 480
 QY 1815 GGGAAAAACGAATTCCTCACTCGTTCGACTTCCATATATGGTTGAGCATACAGTTGAGGTTA 1874
 Db 479 AGATAAATCAGTGGTAAATGTTCAACCACTAATGAGGAAAAACAAAAAGGTAGATGCCA 420
 QY 1875 AAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATAATATTAAATTTCTTGGT 1934
 Db 419 AAACGTAATCAACTACATTAATGTTAAAGTGTATCAAGAGTAATGATGAAAACTC 360
 QY 1935 TTGATGATCACATACAGCACTCAATGCGCTATA 1970
 Db 359 TTGTTGATAACAATAGTAATTCAAATAAGAAAAATA 324

Search completed: September 30, 2004, 22:20:14
 Job time : 188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 15:51:17 ; Search time 1122 Seconds
(without alignments)
10794.292 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTCGGACTGT.....TAAGTAAGGAATAAATAC 2389

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
18:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2388	100.0	2389	9	US-09-765-272-55
2	2388	100.0	2451	13	US-10-412-862-9
3	2388	100.0	2451	16	US-10-412-850-9
4	2388	100.0	2451	16	US-10-387-783-9
5	2388	100.0	8195	13	US-10-158-844-94
6	1374.2	57.5	2523	10	US-09-884-465A-3
7	1374.2	57.5	2647	10	US-09-769-787-206
8	1003.8	42.0	2481	10	US-09-884-465A-4
9	991	41.5	2531	13	US-10-412-862-11
10	991	41.5	2531	16	US-10-412-850-11
11	991	41.5	2531	16	US-10-387-783-11
12	990	41.4	2531	16	US-10-412-862-5
13	990	41.4	2531	16	US-10-412-850-5
14	990	41.4	2531	16	US-10-387-783-5

15	987.6	41.3	2290	9	US-09-765-272-65	Sequence 65, Appli
16	980.8	41.1	2639	10	US-09-884-465A-5	Sequence 5, Appli
17	653.8	27.4	2359	13	US-10-158-844-243	Sequence 243, App
18	385.4	16.1	1342	9	US-09-765-272-181	Sequence 181, App
19	385.4	16.1	1455	10	US-09-769-787-246	Sequence 246, App
20	385.4	16.1	1455	10	US-09-769-744A-23	Sequence 23, Appli
21	385.4	16.1	1455	13	US-10-412-862-7	Sequence 7, Appli
22	385.4	16.1	1455	16	US-10-412-850-7	Sequence 7, Appli
23	385.4	16.1	1455	16	US-10-387-783-7	Sequence 7, Appli
24	385.4	16.1	3120	10	US-09-884-465A-1	Sequence 1, Appli
25	385.4	16.1	5048	10	US-09-884-465A-2	Sequence 2, Appli
26	385.4	16.1	6867	13	US-10-158-844-192	Sequence 192, App
27	381.4	16.0	973	13	US-10-158-844-355	Sequence 355, App
28	347	14.5	2528	10	US-09-884-465A-9	Sequence 9, Appli
29	243.4	10.2	1684	13	US-10-158-844-258	Sequence 258, App
30	211.4	8.8	1146	10	US-09-769-736-23	Sequence 23, Appli
31	211.4	8.8	2469	10	US-09-769-736-17	Sequence 17, Appli
32	209.8	8.8	5215	10	US-09-252-088-13	Sequence 13, Appli
33	209.8	8.8	5215	16	US-10-340-792-13	Sequence 13, Appli
34	201.2	8.4	816	16	US-10-324-143-4	Sequence 4, Appli
35	201.2	8.4	816	16	US-10-324-143-5	Sequence 5, Appli
36	201.2	8.4	819	16	US-09-884-465A-257	Sequence 257, App
37	201.2	8.4	819	16	US-10-324-143-10	Sequence 10, Appli
38	200.8	8.4	2721	16	US-10-324-143-6	Sequence 6, Appli
39	117.4	4.9	1455	10	US-09-769-736-71	Sequence 71, Appli
40	109	4.6	841	9	US-09-452-599-34	Sequence 34, Appli
41	109	4.6	841	15	US-10-121-120-34	Sequence 34, Appli
42	52.8	2.2	1168	16	US-10-292-798-1825	Sequence 1825, Ap
43	51.6	2.2	11076	13	US-10-282-122A-35051	Sequence 35051, A
44	51.2	2.1	1168	15	US-10-017-161-2179	Sequence 2179, Ap
45	50	2.1	573	15	US-10-029-386-25433	Sequence 25433, A
46	49.8	2.1	1959	9	US-09-864-761-4012	Sequence 4012, Ap
47	46.6	2.0	506	15	US-10-029-386-20619	Sequence 20619, A
48	45.4	1.9	58985	10	US-09-901-153-3	Sequence 3, Appli
49	45.4	1.9	143601	10	US-09-855-824-3	Sequence 3, Appli
50	44.6	1.9	423	9	US-09-864-761-18355	Sequence 18355, A
51	44.6	1.9	487	9	US-09-864-761-18355	Sequence 18355, A
52	44.6	1.9	3673778	15	US-10-312-841-2	Sequence 1597, Ap
53	44	1.8	4997	13	US-10-282-122A-35506	Sequence 35506, A
54	43.6	1.8	766	9	US-09-864-761-19608	Sequence 19608, A
55	43.6	1.8	1944	9	US-09-864-761-2825	Sequence 2825, Ap
56	43.4	1.8	1115	13	US-10-425-114-34866	Sequence 34866, A
57	43	1.8	538	15	US-10-029-386-20406	Sequence 20406, A
58	43	1.8	31124	13	US-10-087-192-463	Sequence 463, App
59	42.8	1.8	49753	13	US-10-087-192-1855	Sequence 1855, Ap
60	42.6	1.8	785	15	US-10-029-386-22627	Sequence 22627, A
61	42.6	1.8	684187	17	US-10-367-094-71	Sequence 71, Appli
62	42.4	1.8	660	9	US-09-864-761-19488	Sequence 19488, A
63	42.4	1.8	962	9	US-09-864-761-2772	Sequence 2772, Ap
64	42.4	1.8	4047	9	US-09-815-242-4843	Sequence 4843, Ap
65	42.4	1.8	4050	9	US-09-815-242-9039	Sequence 9039, Ap
66	42.4	1.8	4358	8	US-08-781-986A-454	Sequence 454, App
67	42.4	1.8	4358	13	US-10-329-624-454	Sequence 454, App
68	42.2	1.8	635	13	US-10-027-632-269927	Sequence 269927, A
69	42.2	1.8	635	16	US-10-027-632-269927	Sequence 269927, A
70	42	1.8	305	9	US-09-864-761-19262	Sequence 19262, A
71	42	1.8	496	9	US-09-864-761-2534	Sequence 2534, Ap
72	42	1.8	745	17	US-10-437-963-6282	Sequence 6282, Ap
73	42	1.8	3532	15	US-10-017-161-1913	Sequence 1913, Ap
74	42	1.8	3532	16	US-10-292-798-1569	Sequence 1569, Ap
75	42	1.8	7108	17	US-10-257-166-77	Sequence 77, Appli
76	41.8	1.7	528	15	US-10-029-386-5350	Sequence 5350, Ap
77	41.8	1.7	1010	15	US-10-029-386-24901	Sequence 24901, A
78	41.8	1.7	3673778	15	US-10-312-841-1	Sequence 1, Appli
79	41.6	1.7	300	13	US-10-076-555-378	Sequence 378, App
80	41.6	1.7	413	15	US-10-029-386-14986	Sequence 14986, A
81	41.6	1.7	571	15	US-10-029-386-1284	Sequence 1284, Ap
82	41.6	1.7	611	9	US-09-864-864-233	Sequence 233, App
83	41.6	1.7	1845	9	US-09-864-864-313	Sequence 313, App
84	41.6	1.7	1860	9	US-09-919-497-50	Sequence 50, Appli
85	41.6	1.7	1860	13	US-10-342-887-1928	Sequence 1928, Ap
86	41.6	1.7	1860	13	US-10-172-118-1928	Sequence 1928, Ap
87	41.6	1.7	1860	17	US-10-761-169-8	Sequence 8, Appli

Qy	1381	TAATTCGANTTCCAAGCCTTAGACAAATTAATAGAACGCTTGAAATGATGAATCGACTAA	1440
Db	1381	TAATTCGANTTCCAAGCCTTAGACAAATTAATAGAACGCTTGAAATGATGAATCGACTAA	1440
Qy	1441	TAAAGAAAATTTGCTAGATGATTTATTGGGCATTCCTAGACCAATTAACCCATCCAGAGCG	1500
Db	1441	TAAAGAAAATTTGCTAGATGATTTATTGGGCATTCCTAGACCAATTAACCCATCCAGAGCG	1500
Qy	1501	ACTTTGGCAAAACCAATTTCTCAAAATTGAGTATACHTGAAGAGCAAGCTTCGTATTCCTCAATT	1560
Db	1501	ACTTTGGCAAAACCAATTTCTCAAAATTGAGTATACHTGAAGAGCAAGCTTCGTATTCCTCAATT	1560
Qy	1561	AGCTGATAAGTATACAACGTCAGATGTTTACATTTTGTAGAAATCATGATATAATCACTGTA	1620
Db	1561	AGCTGATAAGTATACAACGTCAGATGTTTACATTTTGTAGAAATCATGATATAATCACTGTA	1620
Qy	1621	TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAGAGATG	1680
Db	1621	TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAGAGATG	1680
Qy	1681	CTTTCTGTGATAAGGAAAAAGTTGCGCTCAAGCCTATACATAAGAAAAAGTATCCTTACC	1740
Db	1681	CTTTCTGTGATAAGGAAAAAGTTGCGCTCAAGCCTATACATAAGAAAAAGTATCCTTACC	1740
Qy	1741	TCCATCTCCAGACGCAGATGTTTAAAGCAATCCAACTGGAGATGTCAGCAGCTATTTTA	1800
Db	1741	TCCATCTCCAGACGCAGATGTTTAAAGCAATCCAACTGGAGATGTCAGCAGCTATTTTA	1800
Qy	1801	CAATCGTGTGAAAGGGGAAAAACGAATTCACATTCGTTCCGATTTCCATATATGTTTGAGCA	1860
Db	1801	CAATCGTGTGAAAGGGGAAAAACGAATTCACATTCGTTCCGATTTCCATATATGTTTGAGCA	1860
Qy	1861	TACAGTTGAGTTTAAAAACGGTAAATTGATTTATTCCTCATAAAGGATCATTTACCATAATAT	1920
Db	1861	TACAGTTGAGTTTAAAAACGGTAAATTGATTTATTCCTCATAAAGGATCATTTACCATAATAT	1920
Qy	1921	TAAATTTGCTTGTTGTTGATGATCACACATACAAGCTCCAAATGGCTATACCTTGGAAAGA	1980
Db	1921	TAAATTTGCTTGTTGTTGATGATCACACATACAAGCTCCAAATGGCTATACCTTGGAAAGA	1980
Qy	1981	TTTGTGTCGACGATTAAGTACTACGTAGAACACCTCGTACGAACTGTCATTTCAATCA	2040
Db	1981	TTTGTGTCGACGATTAAGTACTACGTAGAACACCTCGTACGAACTGTCATTTCAATCA	2040
Qy	2041	TGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA	2100
Db	2041	TGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA	2100
Qy	2101	TAAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	2101	TAAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Qy	2161	TCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTGCAGAA	2220
Db	2161	TCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTGCAGAA	2220
Qy	2221	AGTAAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA	2280
Db	2221	AGTAAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA	2280
Qy	2281	TTTGAATCTTCAAAATTATGGATAACAATAGTATCATGGCAGAAAGCAGAAAAAATTAATTC	2340
Db	2281	TTTGAATCTTCAAAATTATGGATAACAATAGTATCATGGCAGAAAGCAGAAAAAATTAATTC	2340
Qy	2341	GTGTGTTTAAAGGAAGTATCCTTCATCTGTAAGTAAGGAAAAAATAAAC	2399
Db	2341	GTGTGTTTAAAGGAAGTATCCTTCATCTGTAAGTAAGGAAAAAATAAAC	2399

RESULT 2

RESULTS 2
US-10-412-862-9

; Sequence 9, Application US/10412862

661 TTCAAGAACCTATCGCGCAAAATAGCGATAACATCTTCAAGAAACAACTGGGTACCTTC 720
 720 TTCAAGAACCTATCGCGCAAAATAGCGATAACATCTTCAAGAAACAACTGGGTACCTTC 779
 721 TGTAAAGCAATCCAGGAACTACAAATCTTCAAGAAACAACTGGGTACCTTC 780
 780 TGTAAAGCAATCCAGGAACTACAAATCTTCAAGAAACAACTGGGTACCTTC 839
 781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACAGCTCTTCAAACTGGGTAC 840
 840 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACAGCTCTTCAAACTGGGTAC 899
 841 TCAACGACATGTAGATCTGATGGCTTGTCTTGTATCCAGCAAAATCAAACTGGGTAC 900
 900 TCAACGACATGTAGATCTGATGGCTTGTCTTGTATCCAGCAAAATCAAACTGGGTAC 959
 901 AGCTAGAGGTGTGACGTGGCAACACGAGATCATTACCACTTCACTCCCTTACTCTCAAT 960
 960 AGCTAGAGGTGTGACGTGGCAACACGAGATCATTACCACTTCACTCCCTTACTCTCAAT 1019
 961 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTTCGTTATCGTTCAAACTGG 1020
 1020 GTCTGAATTTGAAGAACGAATCGCTCGTATTTATTTCCCTTTCGTTATCGTTCAAACTGG 1079
 1021 GGTACCAAGTTCAGGCAAGAAACAAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
 1080 GGTACCAAGTTCAGGCAAGAAACAAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139
 1081 CCCGCAACCTCCACCAATCTTAAATAGATCTCAAAATCTTCTTGTGTAGTACGTGGT 1140
 1140 CCCGCAACCTCCACCAATCTTAAATAGATCTCAAAATCTTCTTGTGTAGTACGTGGT 1199
 1141 ACAGAAAGTTCAGGCAAGAAATGATTTTCGAAGAAAGGCAATCTCTCGTTATGTTCTTGC 1200
 1200 ACAGAAAGTTCAGGCAAGAAATGATTTTCGAAGAAAGGCAATCTCTCGTTATGTTCTTGC 1259
 1201 GAAAGATTTACCACTGAACTGTTTAAATCTTGAAGCAAGTATCAAAACCAAGAGAG 1260
 1260 GAAAGATTTACCACTGAACTGTTTAAATCTTGAAGCAAGTATCAAAACCAAGAGAG 1319
 1261 TGTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTCTCTCGTACCAAGAAATTTTA 1320
 1320 TGTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTCTCTCGTACCAAGAAATTTTA 1379
 1321 TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCTTGTGTTGNAATTAAGGTCG 1380
 1380 TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCTTGTGTTGNAATTAAGGTCG 1439
 1381 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAACCTTGAATGATGATCACTAA 1440
 1440 TAAATCTGATTTCCAGCCTTAGACAAATTTATGAACCTTGAATGATGATCACTAA 1499
 1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATCTTAGCAACAAATACCCATCCAGAGCG 1500
 1500 TAAAGAAATTTGGTAGATGATTTATTTGGCAATCTTAGCAACAAATACCCATCCAGAGCG 1559
 1501 ACTTGGCAACCAATCTTCAATTTGATGATGATTTATTTGGCAATCTTAGCAACAAAT 1560
 1560 ACTTGGCAACCAATCTTCAATTTGATGATGATTTATTTGGCAATCTTAGCAACAAAT 1619
 1561 ACTTGGCAACCAATCTTCAATTTGATGATGATTTATTTGGCAATCTTAGCAACAAAT 1620
 1620 ACTTGGCAACCAATCTTCAATTTGATGATGATTTATTTGGCAATCTTAGCAACAAAT 1679
 1621 TGAAGGAGATGATATGTAACCGCTCATATGGGCCATAGTCACTGGATTGGAAGATAG 1680
 1680 TGAAGGAGATGATATGTAACCGCTCATATGGGCCATAGTCACTGGATTGGAAGATAG 1739
 1681 CCTTCTGATAGGAAAGTTGACGCTCAAGCTATCTAAGAAAGAGGTATCTTACC 1740
 1740 CCTTCTGATAGGAAAGTTGACGCTCAAGCTATCTAAGAAAGAGGTATCTTACC 1799

1741 TCCATCTCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGGACAGCTATTTTA 1800
 1800 TCCATCTCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGGACAGCTATTTTA 1859
 1801 CAATCTGTTGAAAGGGGAAAGCAATTCACCTCGTTCCAGCTTCCATATATATGTTGAGCA 1860
 1860 CAATCTGTTGAAAGGGGAAAGCAATTCACCTCGTTCCAGCTTCCATATATATGTTGAGCA 1919
 1861 TACAGTTGAGTTTAAAGCGGTAAATTTGATTTATTTCTCTCAATAGGATCATTTACCATAAT 1920
 1920 TACAGTTGAGTTTAAAGCGGTAAATTTGATTTATTTCTCTCAATAGGATCATTTACCATAAT 1979
 1921 TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTTATACCTTTGGA 1980
 1980 TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTTATACCTTTGGA 2039
 1981 TTTGTTTGGCAACCATTAAGTACTACGTAGAACACCTCGACAAAGCTCCATCTTATGA 2040
 2040 TTTGTTTGGCAACCATTAAGTACTACGTAGAACACCTCGACAAAGCTCCATCTTATGA 2099
 2041 TGGATGGGGCAATGCCAGTGGAGCATGTTTGGCAAGAAAGCAACAGTGAAGATCCAAA 2100
 2100 TGGATGGGGCAATGCCAGTGGAGCATGTTTGGCAAGAAAGCAACAGTGAAGATCCAAA 2159
 2101 TAAGAACTTCAAGCGGATGAAGCCAGTGAAGAAACCTCTGAGCCAGAGTCCC 2160
 2160 TAAGAACTTCAAGCGGATGAAGCCAGTGAAGAAACCTCTGAGCCAGAGTCCC 2219
 2161 TCAAGTAGAGCTGAAAGAGTGAAGCCCACTCAAGAAAGCAAGTTCCTTTGGGAA 2220
 2220 TCAAGTAGAGCTGAAAGAGTGAAGCCCACTCAAGAAAGCAAGTTCCTTTGGGAA 2279
 2221 AGTAACGATTTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
 2280 AGTAACGATTTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2339
 2281 TTTGACTCTTCAAAATTTATGATTAAGTGAAGCCCACTCAAGAAAGCAAGTTCCTTTGG 2340
 2340 TTTGACTCTTCAAAATTTATGATTAAGTGAAGCCCACTCAAGAAAGCAAGTTCCTTTGG 2399
 2341 GTTCTTAAAGGAGTATCTTCTCATCTGTAAGTAAAGGAAATTAAC 2389
 2400 GTTCTTAAAGGAGTATCTTCTCATCTGTAAGTAAAGGAAATTAAC 2448

RESULT 3
 US-10-412-850-9
 ; Sequence 9, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 9
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 US-10-412-850-9

Query Match		100.0%;	Score 2388;	DB 16;	Length 2451;		
Best Local Similarity		100.0%;	Pred. No. 0;				
Matches 2389;		Conservative	0;	Mismatches	0;	Indels	Gaps
QY	1	TTCTTACGAGTTGGACGTGATATCAAGCTAGAACGGTTAAGGAAATATATCGTTTCTTA	60				
Db	60	TTCTTACGAGTTGGACGTGATATCAAGCTAGAACGGTTAAGGAAATATATCGTTTCTTA	119				
QY	61	TATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120				
Db	120	TATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	179				
QY	121	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGGCTATGTAC	180				
Db	180	GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGGCTATGTAC	239				
QY	181	TTACATGCGGACCACTATCATATTAATTAACAATGTTAAGGTTCTTATGACGCTATCATCAG	240				
Db	240	TTACATGCGGACCACTATCATATTAATTAACAATGTTAAGGTTCTTATGACGCTATCATCAG	299				
QY	241	TGAAGAAATTAATCAAGGTAGATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC	360				
Db	360	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGC	419				
QY	361	TGCCCCAGGGATAACGTCGTCGTAACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	420				
Db	420	TGCCCCAGGGATAACGTCGTCGTAACAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	479				
QY	421	TCAACATCGTGAAGGTGGAACTCAAGAAACGATGTCGTTGTCCTTGCCTTGCACGTTGCGCA	480				
Db	480	TCAACATCGTGAAGGTGGAACTCAAGAAACGATGTCGTTGTCCTTGCCTTGCACGTTGCGCA	539				
QY	481	AGGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	540				
Db	540	AGGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	599				
QY	541	TGTTGATGCTTATATCGTTCTCATGAGATCATTAACATTAATTCCTAAGAAATCAAGTT	600				
Db	600	TGTTGATGCTTATATCGTTCTCATGAGATCATTAACATTAATTCCTAAGAAATCAAGTT	659				
QY	601	ATCAGTAGCGAGTTGGCTGCTGCAGAACGCTTCTTATCTGTCGAGGAAATCTGTCAA	660				
Db	660	ATCAGTAGCGAGTTGGCTGCTGCAGAACGCTTCTTATCTGTCGAGGAAATCTGTCAA	719				
QY	661	TTCAAGAACCTATGCGCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC	720				
Db	720	TTCAAGAACCTATGCGCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTC	779				
QY	721	TGTAAGCAATCCAGGAACCTACAAATACAAACAGCAACCAAGCAACACTAACAGTCA	780				
Db	780	TGTAAGCAATCCAGGAACCTACAAATACAAACAGCAACCAAGCAACACTAACAGTCA	839				
QY	781	AGCAAGTCAAAAGTAATGACATTTAGTAGTCTCTTGAACAGCTCTACAAATCCCTTTGAG	840				
Db	840	AGCAAGTCAAAAGTAATGACATTTAGTAGTCTCTTGAACAGCTCTACAAATCCCTTTGAG	899				
QY	841	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC	900				
Db	900	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAGTCGAAC	959				
QY	901	AGCTAGAGGTGTGAGTGCACAGGAGATCATTAACGATTCCTTCTTCTTCAAT	960				
Db	960	AGCTAGAGGTGTGAGTGCACAGGAGATCATTAACGATTCCTTCTTCTTCAAT	1019				
QY	961	GTCTGAATTTGGAAGCAAGATCGCTCGTATTTATTCCTTCTGTTATCGTTCAACCAATTG	1020				
Db	1020	GTCTGAATTTGGAAGCAAGATCGCTCGTATTTATTCCTTCTGTTATCGTTCAACCAATTG	1079				

QY	1021	GGTACCAGATTCAAGGCCAGAACCAAGTCCACAACTCCGACTCCGGAACCTAGTCCAGG	1080
Db	1080	GGTACCAGATTCAAGGCCAGAACCAAGTCCACAACTCCGACTCCGGAACCTAGTCCAGG	1139
QY	1081	CCCCCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT	1140
Db	1140	CCCCCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCACTGGT	1199
QY	1141	ACGAAAAGTTGGGAAGGATATGTTATTCGAAGAAAAGGCGATCTCTCGTTATGCTTTGC	1200
Db	1200	ACGAAAAGTTGGGAAGGATATGTTATTCGAAGAAAAGGCGATCTCTCGTTATGCTTTGC	1259
QY	1201	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG	1260
Db	1260	GAAAGATTTACCATCTGAAACTGTTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG	1319
QY	1261	TGTTTTCACACACTTTAACTGCTTAAAGAAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1320
Db	1320	TGTTTTCACACACTTTAACTGCTTAAAGAAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1379
QY	1321	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATTAAGGCTCG	1380
Db	1380	TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATTAAGGCTCG	1439
QY	1381	TAATTTCTGATTTCCAGCCCTTAGACAAATTTATTAGAACGCTTGAATGATGAAATCGACTAA	1440
Db	1440	TAATTTCTGATTTCCAGCCCTTAGACAAATTTATTAGAACGCTTGAATGATGAAATCGACTAA	1499
QY	1441	TAAAGAAAATTTGGTAGATGATTTATTGGCATTTCTTAGACCAATTAACCCATCCAGAGCG	1500
Db	1500	TAAAGAAAATTTGGTAGATGATTTATTGGCATTTCTTAGACCAATTAACCCATCCAGAGCG	1559
QY	1501	ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT	1560
Db	1560	ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAAGTTTCGTTATGCTCAAT	1619
QY	1561	AGCTGATAAGTATACAAACGTCAGATGGTTTACATTTTGTGATGAACATGATATATCATGTGA	1620
Db	1620	AGCTGATAAGTATACAAACGTCAGATGGTTTACATTTTGTGATGAACATGATATATCATGTGA	1679
QY	1621	TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGATTTGGAAGAAAGTAG	1680
Db	1680	TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGATTTGGAAGAAAGTAG	1739
QY	1681	CTTTTCTGATAAGGAAAAGTTGAGCTCAAGCTATATACTAAAGAAAAGGATTCCTACC	1740
Db	1740	CTTTTCTGATAAGGAAAAGTTGAGCTCAAGCTATATACTAAAGAAAAGGATTCCTACC	1799
QY	1741	TCCATCTCCAGACGAGATGTTTAAAGCAATCCACTCGCTTCCATATATATGTTGAGCA	1800
Db	1800	TCCATCTCCAGACGAGATGTTTAAAGCAATCCACTCGCTTCCATATATATGTTGAGCA	1859
QY	1801	CAATCGTGTGAAAGGGGAAAAACGAAATCCACTCGCTTCCATATATATGTTGAGCA	1860
Db	1860	CAATCGTGTGAAAGGGGAAAAACGAAATCCACTCGCTTCCATATATATGTTGAGCA	1919
QY	1861	TACAGTGTGAGTTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTAACATAATAT	1920
Db	1920	TACAGTGTGAGTTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTAACATAATAT	1979
QY	1921	TAAATTTCTGTTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGGAGA	1980
Db	1980	TAAATTTCTGTTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGGAGA	2039
QY	1981	TTTGTTCGACGATTAAGTACTAGTAGAACACCCCTGACCAACCTCCACATTTCTAATGA	2040
Db	2040	TTTGTTCGACGATTAAGTACTAGTAGAACACCCCTGACCAACCTCCACATTTCTAATGA	2099
QY	2041	TGATGGGGCAATGCGCTGAGCATGTTTAGGCAAGAAAGCAACAGTGAAGATCCAAA	2100
Db	2100	TGATGGGGCAATGCGCTGAGCATGTTTAGGCAAGAAAGCAACAGTGAAGATCCAAA	2159
QY	2101	TAAGAACTTCAAGCGGATGAAGGCCAGTAGAGGAAACACCTGCTGAGCCAGGAAGTCCC	2160


```

1440 TAATCTGATTTCCAGCCCTTAGACAAATTAATAGAACGCTTGAATGATGAATCGACTAA 1499
1441 TAAAGAAAATTTGGTAGATGATTTATTGGCATTCTTAGCACCATAATACCATCCAGAGCG 1500
1500 TAAAGAAAATTTGGTAGATGATTTATTGGCATTCTTAGCACCATAATACCATCCAGAGCG 1559
1501 ACTTGGCAACCAAAATTTCAAATTTAGATGATGATGATGATGATGATGATGATGATGATG 1560
1560 ACTTGGCAACCAAAATTTCAAATTTAGATGATGATGATGATGATGATGATGATGATGATG 1619
1561 AGCTGATAGTATCAACAGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
1620 AGCTGATAGTATCAACAGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1679
1621 TGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
1680 TGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739
1681 CCTTCTGATAGGAAAAGTTGAGCTTCAAGCTTATCTAAAGAAAAGGTTATCTTACC 1740
1740 CCTTCTGATAGGAAAAGTTGAGCTTCAAGCTTATCTAAAGAAAAGGTTATCTTACC 1799
1741 TCCATCTCCAGACCCAGATGTTAAAGCAAAATCCAACTGCGATAGTGCAGAGCTATTTA 1800
1800 TCCATCTCCAGACCCAGATGTTAAAGCAAAATCCAACTGCGATAGTGCAGAGCTATTTA 1859
1801 CAATCTGTTGAAAGGGGAAAACGAATTTCCACTCGTTCGATCTCCATATATGTTGAGCA 1860
1860 CAATCTGTTGAAAGGGGAAAACGAATTTCCACTCGTTCGATCTCCATATATGTTGAGCA 1919
1861 TACAGTTGAGGTTAAAGAACGGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 1920
1920 TACAGTTGAGGTTAAAGAACGGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 1979
1921 TAAATTTGCTTGGTTTTCATGATCACAATCAACAGCTTCCAAATGCTTATACCTATAT 1980
1980 TAAATTTGCTTGGTTTTCATGATCACAATCAACAGCTTCCAAATGCTTATACCTATAT 2039
1981 TTTGTTTGGACGATTAAGTACTACGTAGAACACCTTCCAGCAAGCTTCCAAATGCTTAT 2040
2040 TTTGTTTGGACGATTAAGTACTACGTAGAACACCTTCCAGCAAGCTTCCAAATGCTTAT 2099
2041 TGGATGGGCAATGCGAGTGAGATGATGTTAGGCAAGAACCCAGTGAAGATCCAAA 2100
2100 TGGATGGGCAATGCGAGTGAGATGATGTTAGGCAAGAACCCAGTGAAGATCCAAA 2159
2101 TAAGAACTTCAAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2160
2160 TAAGAACTTCAAAGCGGATGAAGAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2219
2161 TCAAGTAGAGCTGAAAAAGTGAAGCCCACTCAAGAGAGCAGAAAGTTTGTCTCGGAA 2220
2220 TCAAGTAGAGCTGAAAAAGTGAAGCCCACTCAAGAGAGCAGAAAGTTTGTCTCGGAA 2279
2221 AGTAAGGATTTCTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
2280 AGTAAGGATTTCTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2339
2281 TTTGACTCTTCAAAATTTATGATAACAAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340
2340 TTTGACTCTTCAAAATTTATGATAACAAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2399
2341 GTTGTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2389
2400 GTTGTAAAGGAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2448

```

RESULT 5

US-10-158-844-94

; Sequence 94, Application US/10158844

; Publication No. US20040029118A1

; GENERAL INFORMATION:

```

;
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
; US-10-158-844-94

```

```

Query Match 100.0%; Score 2388; DB 13; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTCTATCAAGTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 60
Db 3053 TTCTTACGAGTTGGGACTCTATCAAGTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 3112
QY 61 TATAGATGAAAAACAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120
Db 3113 TATAGATGAAAAACAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 3172
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232
QY 181 TTCAATGCGCAGCACCCTATCATTATTAATCAATGTTAGGTTCCCTTATGACGCTATCATCAG 240
Db 3233 TTCAATGCGCAGCACCCTATCATTATTAATCAATGTTAGGTTCCCTTATGACGCTATCATCAG 3292
QY 241 TGAAGAATTTACTCATGAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 300
Db 3293 TGAAGAATTTACTCATGAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 3352
QY 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 360
Db 3353 GGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 3412
QY 361 TGCCACGCGGATACCGTCCGTACAAAAGAGGAATCAATCGACAAAAACAAGAGCATAG 420
Db 3413 TGCCACGCGGATACCGTCCGTACAAAAGAGGAATCAATCGACAAAAACAAGAGCATAG 3472
QY 421 TCAACATCGTGAAGGTGGAACCTCCAAGAAAAAGATGGTCTGTTGGCCTTGGCAGCTTCCCA 480
Db 3473 TCAACATCGTGAAGGTGGAACCTCCAAGAAAAAGATGGTCTGTTGGCCTTGGCAGCTTCCCA 3532

```

QY	481	AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATATAGGATAC	540	Db	4613	AGCTGATAAGTATACAACTGATGTTTACATTTTGTGAACATGATATATCATGTA	4672
Db	3533	AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATATAGGATAC	3592	QY	1621	TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAGATAG	1680
QY	541	TGATGATCTTATATCGTTCTCATCGAGATCATCAATCAATTCCTTAAGATGAGTT	600	Db	4673	TGAAGGAGATGCATATGTAAACGCTCATATGGCCATAGTCACTGGATTGGAAGATAG	4732
Db	3593	TGATGATCTTATATCGTTCTCATCGAGATCATCAATCAATTCCTTAAGATGAGTT	3652	QY	1681	CTTTCTGATAGGAAAGTTGCGAGCTCAAGCTTAAAGCAAAATCCAATCTGGAGATAGT	1740
QY	601	ATCAGCTAGGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTGGAGAAATCTGTCAA	660	Db	4733	CTTTCTGATAGGAAAGTTGCGAGCTTAAAGCAAAATCCAATCTGGAGATAGT	4792
Db	3653	ATCAGCTAGGAGTTGGCTGCTGAGAAAGCTTCTATCTGGTGGAGAAATCTGTCAA	3712	QY	1741	TCCATCTCCAGACGAGATCTTAAAGCAAAATCCAATCTGGAGATAGTCACTGGAGCA	1800
QY	661	TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGTACCTTC	720	Db	4793	TCCATCTCCAGACGAGATCTTAAAGCAAAATCCAATCTGGAGATAGTCACTGGAGCA	4852
Db	3713	TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGTACCTTC	3772	QY	1801	CAATCTGCTGAAAGGGGAAAGAAAGCAATTCCTGCTTCCGACTTCCATATATGTTGAGCA	1860
QY	721	TGTAAGCAATCCAGAACTACAAATCTAACAAGCAAAACAAACAGCAACACTCAACAGTCA	780	Db	4853	CAATCTGCTGAAAGGGGAAAGAAAGCAATTCCTGCTTCCGACTTCCATATATGTTGAGCA	4912
Db	3773	TGTAAGCAATCCAGAACTACAAATCTAACAAGCAAAACAAACAGCAACACTCAACAGTCA	3832	QY	1861	TACAGTTGAGTTAAAAACCGTAAATTTGATTTATTTCTCATAAGGATCATTAACCATATAT	1920
QY	781	AGCAAGTCAAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840	Db	4913	TACAGTTGAGTTAAAAACCGTAAATTTGATTTATTTCTCATAAGGATCATTAACCATATAT	4972
Db	3833	AGCAAGTCAAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3892	QY	1921	TAAATTTGCTTGGTTTGGATGATCAACATCAAAAGCTCCAAATGGCTATATCTTTGGAAGA	1980
QY	841	TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	900	Db	4973	TAAATTTGCTTGGTTTGGATGATCAACATCAAAAGCTCCAAATGGCTATATCTTTGGAAGA	5032
Db	3893	TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3952	QY	1981	TTTGTGTCGACGATTAAGTACTAGTAGAACAACCTGAGCAACCTGAGCAATCTTAATGA	2040
QY	901	AGCTAGAGGTTGTCAGTGCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAT	960	Db	5033	TTTGTGTCGACGATTAAGTACTAGTAGAACAACCTGAGCAACCTGAGCAATCTTAATGA	5092
Db	3953	AGCTAGAGGTTGTCAGTGCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAT	4012	QY	2041	TGGATGGGCAATGCGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2100
QY	961	GTCGAAATGGAAGAACCAATCGCTGATTAATCCCTTCTGTTATCGTTTCAACCATG	1020	Db	5093	TGGATGGGCAATGCGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	5152
Db	4013	GTCGAAATGGAAGAACCAATCGCTGATTAATCCCTTCTGTTATCGTTTCAACCATG	4072	QY	2101	TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCC	2160
QY	1021	GGTACAGATTCAGGCGAGAACCAACCAAGTCCACCGACTCCGGAACCTAGTCCAGG	1080	Db	5153	TAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAAAGTCC	5212
Db	4073	GGTACAGATTCAGGCGAGAACCAACCAAGTCCACCGACTCCGGAACCTAGTCCAGG	4132	QY	2161	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAAAGTTTCTGCTGCGAA	5272
QY	1081	CCGCAACTGACCAACCAATCTTAAATAGACTCAATTTCTTCTGTTAGTCACTGCT	1140	Db	5213	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAAAGTTTCTGCTGCGAA	5280
Db	4133	CCGCAACTGACCAACCAATCTTAAATAGACTCAATTTCTTCTGTTAGTCACTGCT	4192	QY	5273	AGTAAACGATTCAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATTA	5332
QY	1141	ACGAAAGTTGGGAGGATATGATTCGAAGAAAGGCACTCTCGTTATGCTTTTC	1200	Db	2281	TTTGACTCTTCAAAATTTATGATAAATAGTATCATGGCAGAGCAAGAAATTAATCTTGC	2340
Db	4193	ACGAAAGTTGGGAGGATATGATTCGAAGAAAGGCACTCTCGTTATGCTTTTC	4252	QY	5333	TTTGACTCTTCAAAATTTATGATAAATAGTATCATGGCAGAGCAAGAAATTAATCTTGC	5392
QY	1201	GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG	1260	Db	2341	GTGTTTAAAGGAGTAATCTTCTATCTGTAGTAAGGAAATTAATTAAC	2389
Db	4253	GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG	4312	QY	5393	GTGTTTAAAGGAGTAATCTTCTATCTGTAGTAAGGAAATTAATTAAC	5441
QY	1261	TGTTTCCACACTTTAATCTGTTAAAGAAAGAAATGTTGCTTCTCGTGAACAAATTTTA	1320				
Db	4313	TGTTTCCACACTTTAATCTGTTAAAGAAAGAAATGTTGCTTCTCGTGAACAAATTTTA	4372				
QY	1321	TGTAAGCATATAATCTGTTAACTGAGGCTCATTAAGCCTTCTTGAATTAAGGCTG	1380				
Db	4373	TGTAAGCATATAATCTGTTAACTGAGGCTCATTAAGCCTTCTTGAATTAAGGCTG	4432				
QY	1381	TAATCTGATTTCCAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA	1440				
Db	4433	TAATCTGATTTCCAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA	4492				
QY	1441	TAAAGAAAAATTTGATGATGATTTATTTGGCATCTTAGCACAATTAACCATCCAGAGG	1500				
Db	4493	TAAAGAAAAATTTGATGATGATTTATTTGGCATCTTAGCACAATTAACCATCCAGAGG	4552				
QY	1501	ACTTGGCAACCAATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1560				
Db	4553	ACTTGGCAACCAATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	4612				
QY	1561	AGCTGATAAGTATACAACTGATGGTTTATACATTTTGTGAACATGATATATATCATGTA	1620				

RESULT 6
 US-09-884-465A-3
 : Sequence 3, Application US/09884465A
 : Publication No. US20030077293A1
 : GENERAL INFORMATION:
 : APPLICANT: Shire Biochem, Inc.
 : APPLICANT: Hamel, Josee
 : APPLICANT: Brodeur, Bernard
 : APPLICANT: Martin, Denis
 : APPLICANT: Ouellet, Nathalie
 : APPLICANT: Ouellet, Catherine
 : TITLE OF INVENTION: Streptococcus Antigens
 : FILE REFERENCE: 055190-0044
 : CURRENT APPLICATION NUMBER: US/09/884, 465A
 : CURRENT FILING DATE: 2001-06-20
 : PRIOR APPLICATION NUMBER: 60/212,683
 : PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-03-884-465A-3

Query Match 57.5%; Score 1374.2; DB 10; Length 2523;
 Best Local Similarity 73.9%; Pred. No. 0;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY	1	TTCTTAGAGTTGGAGCTGTATCAAGCTAGACGGTTAAGGAAATATATCTGTTTCCTA	60
DB	60	TGCTTATGAATAGGTTTGCATCAAGCTCAAACTGTAAGAAATATATCTGTTTCCTA	119
QY	61	TATAGATGGAAACCAAGCGCGCAAAACCGAGAAATTTGACTCTGATGAGTTAGCAA	120
DB	120	TATAGATGGAAACCAAGCGCGCAAAACCGAGAAATTTGACTCTGATGAGTTAGCAA	179
QY	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATACAGCAACGAGCTATGTCAC	180
DB	180	GCCTGAAGGAATCAAGCGCGCAAAACCGAGAAATTTGACTCTGATGAGTTAGCAA	239
QY	181	TTCAATGCGGACCACTATCATTAATCAATGATGATGAGTTTCTTATGACGCTATCATCAG	240
DB	240	CTCTCATGAGAGCACTATCATTAATGATGATGAGTTTCTTATGACGCTATCATCAG	299
QY	241	TGAAGATTTACTCATGAAAGATCCAACTATAGCTAAAGATGAGGATATTTGTAATGA	300
DB	300	TGAAGAGCTCTCATGAAAGATCCGAAATTTATCAGTTGAAGGATTCAGACATTTGTAATGA	359
QY	301	GGTCAAGGGTGGATGATGTTCAAGGTAGATGGAATAATCAATCGACAAACCAAGAGCATAG	420
DB	360	AATCAAGGGTGGTATGTTCAATGATGATGGAATAATCAATCGACAAACCAAGAGCATAG	479
QY	361	TGCGCCAGCGGATACCGTCCGTAACAAAGAGGAATCAATCGACAAACCAAGAGCATAG	480
DB	420	AGCTCATGCGGATTAATGTCGTAACAAAGAGGAATCAATCGACAAACCAAGAGCATAG	539
QY	421	TCAACATCGTGAAGGTGGAACTCAAGAAACGATGGTGGTGGTCCCTGGCAAGTTGCGA	540
DB	480	TAGCATCGTGAAGGTGGAACTCAAGAAACGATGGTGGTGGTCCCTGGCAAGTTGCGA	599
QY	481	AGGACGCTATACACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGATAC	600
DB	540	GGGACGCTATACACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGATAC	659
QY	541	TGGTATGCTTATATCGTTCCCTCATGGAGATCAATACCATTAATTCCTAAGAAATGAGTT	660
DB	600	GGGCGATGCTTATATCGTTCCCTCATGGAGATCAATACCATTAATTCCTAAGAAATGAGTT	660
QY	601	ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCTTATCTGGTTCGAGGAAATCTGTCAAA	719
DB	660	ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCTTATCTGGTTCGAGGAAATCTGTCAAA	720
QY	661	TTCAAGAACTTATCGCGGCAAAATAGGATACACTTCAAGAAACAAATCTGGGTACCTTC	779
DB	720	TTTAAAGAACTTATCGCGGCAAAATAGGATACACTTCAAGAAACAAATCTGGGTACCTTC	780
QY	721	TGTAGCAATCCAGGAACTACAAATACCTAACCAAGCAACACAGCAACACTAACAGTCA	839
DB	780	TGTAGCAATCCAGGAACTACAAATACCTAACCAAGCAACACAGCAACACTAACAGTCA	840
QY	781	AGCAAGTCAAGCTAATGATGATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG	899
DB	840	AGCAAGTCAAGCTAATGATGATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG	900
QY	841	TCAACGATGATGATGATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG	959
DB	900	TCAACGATGATGATGATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG	960
QY	901	AGCTAGAGGTGTTGAGTGCCACAGGAGATCAATACCATTTCTATCTCTCAAT	

DB	960	CGCCAGAGGTGTAGCTGTCCCTCATGTGTAACCATTTATCCCTTATGAACAAAT	1019
QY	961	GTCTGAATTTGAAGAAACGAATCGCTCGTATTATTATTTCCCTTCTGTTATCGTTCAAAACATG	1020
DB	1020	GTCTGAATTTGAAGAAACGAATCGCTCGTATTATTATTTCCCTTCTGTTATCGTTCAAAACATG	1079
QY	1021	GGTACCAGATTTCAAGGGCCAGAACCAAGTCCACACCGACTCCGGACCTAGTCCAGG	1080
DB	1080	GGTACCAGATTTCAAGGGCCAGAACCAAGTCCACACCGACTCCGGACCTAGTCCAGG	1139
QY	1081	CCCGCAACCTGTCACCAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT	1128
DB	1140	TCCGCAACCTGTCACCAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT	1199
QY	1129	TAGTCAGCTGTGTACGAAAGTTGGGGAAGGATATGTTTGAAGAAAGGCAATCCAATTTGATGAAATTTGGT	1188
DB	1200	CAAGAAGCTGTTTGAAGAGTTAGCGGATGTTTGTCTTTGAGGAGGATTTGAGTTTCTCG	1259
QY	1189	TTATGCTTTTGGCGAAGATTTTACCATCTGAACTGTTTAAATACTTGAAGCAAGTTTATC	1248
DB	1260	TTATATCCAGCCCAAGAAATCTTTTCAAGCAAGAACAGCAGCGGCAATGATAGCAACTGGC	1319
QY	1249	AAACAGAGAGGTGTTTTCACACACTTTAACTGCTTAAATAAGAAATTTGCTTCTCTGTA	1308
DB	1320	CAAGCAGGAAAGTTTATCTCATAGCTTAGGAGCTTAAAGAACTGACCTCCATCTAGTGA	1379
QY	1309	CCAAGATTTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTG	1368
DB	1380	TCGAGATTTTCAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTTGA	1439
QY	1369	AAATAAGGGTGGTAAATTTCTGAATTTCCAAGCTTTAGACAAATTTATAGAACCTTTGAATGA	1428
DB	1440	TAATAAAGGTTCGCAAGTTGATTTTGAAGCTTTTGAATTAACCTGTTGAAACGACTCAAGGA	1499
QY	1429	TGAATCGACTTAATAAGAAATTTGTTAGATGATTTATTTGGCATTTCTAGACCAATATAC	1488
DB	1500	TGCTCAAGTGATTAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGCTCCGATTCG	1559
QY	1489	CCATCCAGAGCGACTTTGGCAACCAAAATTTCTCAAAATTTAGTATACCTGAAGACCAAGTTTCG	1548
DB	1560	TCATCCAGAGCTTTAGGAAACCAAAATGGCAAAATTAACCTACACTGATGATGAGATTCA	1619
QY	1549	TATGCTCAATTTAGCTGATTAAGTATACAAGTCAAGCTCAGATGTTTACATTTTGTATGAAACATGA	1608
DB	1620	AGTAGCCAAAGTTGGCAGGCAAGTACACACAGAAAGCGTTATATCTTTGATCTCTGTA	1679
QY	1609	TATAATCAGTGATGAGGAGATGTCATATGTAAGCTCATATGCGGCTCATATGGGCTAGTCACTGGAT	1668
DB	1680	TATAACCGATGATGAGGAGATGTCATATGTAAGCTCATATGCGGCTCATATGGGCTAGTCACTGGAT	1739
QY	1669	TGGAAAAGATAGCTTTCTGATAAGGAAAAGTTGCGAGCTCAAGCTCATACCTAATAAGAAA	1728
DB	1740	TAAAAAGATAGTTTGTCTGAAGCTTGAAGAGCGGCGAGCCCAAGCTTTATGCTTAAAGAGAA	1799
QY	1729	AGGTATCTTATCTCCATCTCCAGACCGAGATTTAAAGCAAAATCCAATCTGGAGATAGTGC	1788
DB	1800	AGGTTTGACCCCTCTCTCCAGACCGATCCAGATTTAGGAAATCTAGGCAAAAGGAGC	1859
QY	1789	AGCAGCTTTTACATCGTGTGAAAGGGGAAAGAAAGAAATTCACCTGCTTCCATTA	1848
DB	1860	AGAGCTATCTACAAACCGGCTGAAAGAGAGCTTAAAGAGGTGAGCTTGTATGCTTATGCTT	1919
QY	1849	TATGTTGAGCATACAGTTGAGTTTAAAAACGTAATTTTGAATTTATCTCTCAATAGGATCA	1908
DB	1920	CAATCTCAATATCTGTAGAGTCAAAACGTTGTTTAACTATACCTCATTTATGACCA	1979
QY	1909	TTACCATTAATTTAAATTTGCTTTGATGATCACACATACAAAGCTCCAAATGGGCTA	1968
DB	1980	TTACCATTAATTTAAATTTGCTTTGATGATCACACATACAAAGCTCCAAATGGGCTA	2039
QY	1969	TACCTTGAAGATTTGTTGGAGATTAAGTACTACGTAGAACACCTCGACACGCTCC	2028

Fri Oct 1 15:35:47 2004

Db	2040	TACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTTCGAACATCCAAACGAAACGTC	2099	Db	284	CTCTCATGGAGACCATTTATCAITTAATAATGGCAAGTCCCTTTATGATGCCATCATCAG	343
Qy	2029	ACATTTCTAATGATGGATGGGGCAATGCCAGTAGAGCATGTGTTAGGCGAAGAACCCACAG	2088	Qy	241	TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATGTTAATGA	300
Db	2100	GCATTCAGATAAATGTTTGGTAACGCTAGCGACCATGTTCAAGAGAAACAAAAATGGTCA	2159	Db	344	TGAAGAGCTCTCTCATGAAGATCGGAATTAATCAGTTGAAGGATTCAGACATTTGTCATGA	403
Qy	2089	TGAAGATCCAAATAAGAACTTCAAGCGGATGAAGAG-----	2125	Qy	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Db	2160	AGCTGATACCAATCAAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGA	2219	Db	404	AATCAAGGGTGGTATGTCATTAAGGTAAACGGGTAAATACTATGTTTACCTTAAGGATGC	463
Qy	2126	-----	2125	Qy	361	TGCCACGGCGATTAACGTTCCTACAAAGAGGAATCAATCGACAAAAAACAAGACATAG	420
Db	2220	GGAAGAAACCCCTCGAGNAGAGAACCCACAAAGCGAGAACACAGAGTCTCCAAACCAAC	2279	Db	464	AGCTCATGGCGGATAATGTCGTACAAAGAGGAATCAATCGGCAAAAAACAAGACATAG	523
Qy	2126	-----CAGTAGAGGAAACACCTGCTGAGCCGAGAGTCCCTCAAGTAGAGACTGAAAA	2178	Qy	421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTGTCTTGGCTTGGCACGCTTCGCA	480
Db	2280	AGAGGAACCAAGAGAAATCACCAGAGGAATCAGAGAACTCAGGTGCGAGACTGAAAA	2339	Db	524	TCAGCATCGTGAAGGAGGGACTTCAGCAACGATGGTGGGTAGCCCTTGCACGTTCAACA	583
Qy	2179	AGTAGAGCCCAACTCAAGAGAGCAGAGTCTTTCGTTGCGAAAGTAAACGGATTTCTAGTCT	2238	Qy	481	AGGACGCTATACATACAGATGATGTTTAAATCTTAAATGCTTCTGATATCATAGAGTAC	540
Db	2340	GGTTGAAGAAAACTGAGAGAGGTGAAGATTACTTTGGAATAATCCAGGATCCAAATAT	2399	Db	584	GGGACGCTACACCACAGATGATGTTTATCTTCAATGCAATCTGATATCATCGAGATAC	643
Qy	2239	GAAGGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAATTTGACTCTTCAAAATAT	2298	Qy	541	TGGTATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	600
Db	2400	CAAGTCCCAATGCCAAGAGACTCTCACAGGATTAAGAAATAATTTACTATTTGGCACCCA	2459	Db	644	GGGCGATGCCCTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	703
Qy	2299	GGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTGGTTTAAAGGAGTAA	2358	Qy	601	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCAGGAAATTCGTCAA	660
Db	2460	GGACAACAATACTATTATGGCAGAGCTGAAATAACTATTGGCTTTTAAAGGAGTAA	2519	Db	704	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCGGGAAATTCGTCAA	763
RESULT 7							
US-09-884-465A-4							
; Sequence 4, Application US/09884465A							
; Publication No. US20030077293A1							
; GENERAL INFORMATION:							
; APPLICANT: Shire Biochem, Inc.							
; APPLICANT: Hamel, Josee							
; APPLICANT: Brodeur, Bernard							
; APPLICANT: Martin, Denis							
; APPLICANT: Charland, Nathalie							
; APPLICANT: Ouellet, Catherine							
; TITLE OF INVENTION: Streptococcus Antigens							
; FILE REFERENCE: 055190-0044							
; CURRENT APPLICATION NUMBER: US/09/884,465A							
; CURRENT FILING DATE: 2001-06-20							
; PRIOR APPLICATION NUMBER: 60/212,683							
; PRIOR FILING DATE: 2000-06-20							
; NUMBER OF SEQ ID NOS: 384							
; SOFTWARE: PatentIn version 3.1							
; SEQ ID NO 4							
; LENGTH: 2647							
; TYPE: DNA							
; ORGANISM: Streptococcus pneumoniae							
US-09-884-465A-4							
Query Match 57.5%; Score 1374.2; DB 10; Length 2647;							
Best Local Similarity 73.9%; Pred. No. 0;							
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;							
Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA	60	Qy	1081	CCGCAACCTGCACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGTT	1128
Db	104	TGCTTATGACTAGGTTTGCTCAAGCTCAAGCTGTAAGAAAAATAATCGTGTTCCTA	163	Db	1184	TCGCAACCTGCACCAAAATCTTCAACCACTCCAGCAATCCCAATGATGAGAAATGTT	1243
Qy	61	TATAGATGGAACAAACGAGCGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	120	Qy	1129	TAGTCAGCTGGTACGAAAAAGTTGGGAAGGATATGTTTCGAAGAAAAAGGCAATCTCTCG	1188
Db	164	TATAGATGGAACAAACGAGCGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	223	Db	1244	CAAAGAGCTGTTGCAAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAAATGGAGTTCTCG	1303
Qy	121	GGGTGAGGAATCAATGCTCAGCAAAATCGTCATCAGATAACAGACCAAGGCTATGTCAC	180	Qy	1189	TTATGCTTTTGGAAAGATTTACCATCTGAACCTGTTTAAATCTTTGAAAGCAAGTTATC	1248
Db	224	GGGTGAGGAATCAACGCCCAACAAATCGTCATCAGATTCAGATCAAGTTATGTGAC	283	Db	1304	TTATATCCAGCCCAAGAAATCTTTTCAGCAGAAACAGCAGAGGCAATGATGAGCAATGGC	1363
Qy	181	TTCCATGGCGCACTATCATTTATTAATGTTAAGGTTTCTTATGACGCTATCATCAG	240	Qy	1249	AAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTA	1308

Db	284	CTCTCATGGAGACCATTTATCAITTAATAATGGCAAGTCCCTTTATGATGCCATCATCAG	343	Db	284	CTCTCATGGAGACCATTTATCAITTAATAATGGCAAGTCCCTTTATGATGCCATCATCAG	343
Qy	241	TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATGTTAATGA	300	Qy	241	TGAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATGTTAATGA	300
Db	344	TGAAGAGCTCTCTCATGAAGATCGGAATTAATCAGTTGAAGGATTCAGACATTTGTCATGA	403	Db	344	TGAAGAGCTCTCTCATGAAGATCGGAATTAATCAGTTGAAGGATTCAGACATTTGTCATGA	403
Qy	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360	Qy	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC	360
Db	404	AATCAAGGGTGGTATGTCATTAAGGTAAACGGGTAAATACTATGTTTACCTTAAGGATGC	463	Db	404	AATCAAGGGTGGTATGTCATTAAGGTAAACGGGTAAATACTATGTTTACCTTAAGGATGC	463
Qy	361	TGCCACGGCGATTAACGTTCCTACAAAGAGGAATCAATCGACAAAAAACAAGACATAG	420	Qy	361	TGCCACGGCGATTAACGTTCCTACAAAGAGGAATCAATCGACAAAAAACAAGACATAG	420
Db	464	AGCTCATGGCGGATAATGTCGTACAAAGAGGAATCAATCGGCAAAAAACAAGACATAG	523	Db	464	AGCTCATGGCGGATAATGTCGTACAAAGAGGAATCAATCGGCAAAAAACAAGACATAG	523
Qy	421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTGTCTTGGCTTGGCACGCTTCGCA	480	Qy	421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTGTCTTGGCTTGGCACGCTTCGCA	480
Db	524	TCAGCATCGTGAAGGAGGGACTTCAGCAACGATGGTGGGTAGCCCTTGCACGTTCAACA	583	Db	524	TCAGCATCGTGAAGGAGGGACTTCAGCAACGATGGTGGGTAGCCCTTGCACGTTCAACA	583
Qy	481	AGGACGCTATACATACAGATGATGTTTAAATCTTAAATGCTTCTGATATCATAGAGTAC	540	Qy	481	AGGACGCTATACATACAGATGATGTTTAAATCTTAAATGCTTCTGATATCATAGAGTAC	540
Db	584	GGGACGCTACACCACAGATGATGTTTATCTTCAATGCAATCTGATATCATCGAGATAC	643	Db	584	GGGACGCTACACCACAGATGATGTTTATCTTCAATGCAATCTGATATCATCGAGATAC	643
Qy	541	TGGTATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	600	Qy	541	TGGTATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	600
Db	644	GGGCGATGCCCTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	703	Db	644	GGGCGATGCCCTATATCGTTCCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	703
Qy	601	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCAGGAAATTCGTCAA	660	Qy	601	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCAGGAAATTCGTCAA	660
Db	704	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCGGGAAATTCGTCAA	763	Db	704	ATCAGCTAGCGAGTTGGCTGCTGAGAGGCTTCTCTATCTGGTTCGGGAAATTCGTCAA	763
Qy	661	TTCAAGAACCTTATCGCGACAAAAATAGCGATAACACATTCAGAGCAAACTGGGTACCTTC	720	Qy	661	TTCAAGAACCTTATCGCGACAAAAATAGCGATAACACATTCAGAGCAAACTGGGTACCTTC	720
Db	764	TTTAAAGAACCTTATCGCGACAAAAATAGCGATAACACATTCAGAGCAAACTGGGTACCTTC	823	Db	764	TTTAAAGAACCTTATCGCGACAAAAATAGCGATAACACATTCAGAGCAAACTGGGTACCTTC	823
Qy	721	TGTAAGCAATCCAGAACTTACAAATACTAACAGCAAGCAACCAACACACACTAAACAGTCA	780	Qy	721	TGTAAGCAATCCAGAACTTACAAATACTAACAGCAAGCAACCAACACACACTAAACAGTCA	780
Db	824	TGTAAGCAATCCAGAACTTACAAATACTAACAGCAAGCAACCAACACACACTAAACAGTCA	883	Db	824	TGTAAGCAATCCAGAACTTACAAATACTAACAGCAAGCAACCAACACACACTAAACAGTCA	883
Qy	781	AGCAAGTCAAAAGTAAATGACATTTGATGATCTCTTGAACAGCTCTCAAACTGCCCTTTGAG	840	Qy	781	AGCAAGTCAAAAGTAAATGACATTTGATGATCTCTTGAACAGCTCTCAAACTGCCCTTTGAG	840
Db	884	AGCAAGTCAAAAGTAAATGACATTTGATGATCTCTTGAACAGCTCTCAAACTGCCCTTTGAG	943	Db	884	AGCAAGTCAAAAGTAAATGACATTTGATGATCTCTTGAACAGCTCTCAAACTGCCCTTTGAG	943
Qy	841	TCACGCAATGTAGAAATCTGATGGCTTGTCTTTCATCCAGCACAACCAACCAAGTCGAAC	900	Qy	841	TCACGCAATGTAGAAATCTGATGGCTTGTCTTTCATCCAGCACAACCAACCAAGTCGAAC	900
Db	944	TCACGCAATGTAGAAATCTGATGGCTTGTCTTTCATCCAGCACAACCAACCAAGTCGAAC	1003	Db	944	TCACGCAATGTAGAAATCTGATGGCTTGTCTTTCATCCAGCACAACCAACCAAGTCGAAC	1003
Qy	901	AGCTAGAGGTGTTGAGTGCCACACAGGAGATCATTAACCACTTCACTCCCTTACTCTCAAA	960	Qy	901	AGCTAGAGGTGTTGAGTGCCACACAGGAGATCATTAACCACTTCACTCCCTTACTCTCAAA	960
Db	1004	CGCCAGAGGTGATGTCCTCTCATGGTAAACCATTAACCACTTATCCCTTTATGACAAAT	1063	Db	1004	CGCCAGAGGTGATGTCCTCTCATGGTAAACCATTAACCACTTATCCCTTTATGACAAAT	1063
Qy	961	GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTTCGTTATCGTTTCAAAACATTG	1020	Qy	961	GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTTCGTTATCGTTTCAAAACATTG	1020
Db	1064	GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTTCGTTATCGTTTCAAAACATTG	1123	Db	1064	GTCTGAATTTGGAAGAACCAATCGCTCGTATTAATCCCTTTCGTTATCGTTTCAAAACATTG	1123
Qy	1021	GGTACAGATTCAGAGCCAGAAACACCAAGTCCCAACCGACTCCCGAACCCTAGTCCAGS	1080	Qy	1021	GGTACAGATTCAGAGCCAGAAACACCAAGTCCCAACCGACTCCCGAACCCTAGTCCAGS	1080
Db	1124	GGTACAGATTCAGAGCCAGAAACACCAAGTCCCAACCGACTCCCGAACCCTAGTCCAGS	1183	Db	1124	GGTACAGATTCAGAGCCAGAAACACCAAGTCCCAACCGACTCCCGAACCCTAGTCCAGS	1183
Qy	1081	CCGCAACCTGCACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGTT	1128	Qy	1081	CCGCAACCTGCACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGTT	1128
Db	1184	TCGCAACCTGCACCAAAATCTTCAACCACTCCAGCAATCCCAATGATGAGAAATGTT	1243	Db	1184	TCGCAACCTGCACCAAAATCTTCAACCACTCCAGCAATCCCAATGATGAGAAATGTT	1243
Qy	1129	TAGTCAGCTGGTACGAAAAAGTTGGGAAGGATATGTTTCGAAGAAAAAGGCAATCTCTCG	1188	Qy	1129	TAGTCAGCTGGTACGAAAAAGTTGGGAAGGATATGTTTCGAAGAAAAAGGCAATCTCTCG	1188
Db	1244	CAAAGAGCTGTTGCAAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAAATGGAGTTCTCG	1303	Db	1244	CAAAGAGCTGTTGCAAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAAATGGAGTTCTCG	1303
Qy	1189	TTATGCTTTTGGAAAGATTTACCATCTGAACCTGTTTAAATCTTTGAAAGCAAGTTATC	1248	Qy	1189	TTATGCTTTTGGAAAGATTTACCATCTGAACCTGTTTAAATCTTTGAAAGCAAGTTATC	1248
Db	1304	TTATATCCAGCCCAAGAAATCTTTTCAGCAGAAACAGCAGAGGCAATGATGAGCAATGGC	1363	Db	1304	TTATATCCAGCCCAAGAAATCTTTTCAGCAGAAACAGCAGAGGCAATGATGAGCAATGGC	1363
Qy	1249	AAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTA	1308	Qy	1249	AAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTA	1308

Db 1364 CAAGCAGGAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTGACCTCCCATCTAGTGA 1423
 Qy 1309 CCAAGAAATTTTATGATAAAGCATATAATCTGTGTTAACTGAGGCTCATAAAGCCTTGTTGN 1368
 Db 1424 TCGAGAAATTTTACATAAAGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTTGA 1483
 Qy 1369 AAATAAGGGTCGTAAATCTGATTTCCAAAGCCCTTACACAAATTTATAGAACGCTTGAATGA 1428
 Db 1484 TAATAAAGGTCGCAAGTTGATTTTGGGCTTTGGATAACCTGTTGGAAACGACTCAAGGA 1543
 Qy 1429 TGAATCGACTAATAAAGAAATTTGGTAGATGATTTATGGCATTCCTAGCACCATAATPAC 1488
 Db 1544 TGTCTCAAGTATAAAGTCAAGTTAGTGGATGATTTCTTGCCCTTCTTAGCTCCGATTCG 1603
 Qy 1489 CCATCCAGAGCGACTTGGCAAAACCAATTTCTCAAAATTTAGTATACCTAGAGCAAGTTTCG 1548
 Db 1604 TCATCCAGAAGTTTAGGAACCAATGCGCAATTTACCTACACTGATGATGAGATTCGA 1663
 Qy 1549 TATTTGCTCAATTAGCTGATTAAGTATACAAAGTTCAGATGGTTTACATTTTGTATGACATGA 1608
 Db 1664 AGTAGCCAAAGTTGGCAGGCAAGTACACACAGAGACGGTTATATCTTTGATCTCTCGTA 1723
 Qy 1609 TATAATCAGTATGAGGAGATGATGATGTAACGCTCATATGGGCGCATAGTCACTCGAT 1668
 Db 1724 TATAACCGTATGAGGGGGATGCTATGTAACCTCCACATATGACCCATAGCCACTCGAT 1783
 Qy 1669 TGGAAAGATAGGCTTTCTGATAAGGAAAGTTGACAGCTCAAGCCTATATCTAAAGAAA 1728
 Db 1784 TAAAAAAGATGTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAAGAGAA 1843
 Qy 1729 AGGTATCTACCTCCATCTCCAGGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGC 1788
 Db 1844 AGGTTTGAACCCCTCTTCCAGACAGACCATCAGGATTCAGGAATTAATGAGCAAAAGGAGC 1903
 Qy 1789 AGCAGCTATTTCACATCTGTGTAAGGGAAGAAACGATTCCTACTGCTTGCATCTCCATA 1848
 Db 1904 AGAAGCTATCTACACCGGTGAAGGAGCTAAGAGGTCGCACTTATGCTATGCTGCTTA 1963
 Qy 1849 TATGGTTGAGCATACAGTTGAGGTTAAAAACGGTAAATTTGATTTATTCCTCATAGATCA 1908
 Db 1964 CAATCTTCAATATCTAGTAGAGTCAAAACGGTAGTTTAAATCATCTCTCATTTATGACCA 2023
 Qy 1909 TTACCTAATATTAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTA 1968
 Db 2024 TTACCTAATATTAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTA 2083
 Qy 1969 TACCTTGAAGATTTGTTGCGAGGATTAAGTACTAGTAGAAGACCTGACGAAAGTCC 2028
 Db 2084 TACTCTGAGGATCTTTTGGGACTGTCAAGTACTATGTCGAACATCCAAACGAAAGTCC 2143
 Qy 2029 ACATTTCTAATGATGGGCAATGCCAGTGAAGTATGTTAGGCAAGAAAGACCAAG 2088
 Db 2144 GCATTCAGATAATGTTTGGTTAAACGCTAGGACCATGTTCAAGAAACAAATGGTCA 2203
 Qy 2089 TGAAGTCCAAATAGAACTTCAAAGCGGATGAAGAG-----2125
 Db 2204 AGCTGATACCAATCAACCGGAAACCAAGCGAGGAGAAACCTCAGACAGAAACCTTGA 2263
 Qy 2126 -----2125
 Db 2264 GGAAGAAACCCCTCGAGAGAGAAACCAACAAAGCGAGAAACAGAGTCTTCCAAAACCAAC 2323
 Qy 2126 -----CCAGTAGAGGAAACACCTGCTGAGCGAGAGTCCCTCAAGTAGAGACTGAAAA 2178
 Db 2324 AGAGGAAACCAAGAGAAATCACAGAGGAATCAGAGAACCTCAGGTCGAGACTGAAAA 2383
 Qy 2179 AGTAGAGCCCACTCAAGAGAGCAAGTTTGTCTTGGAAAGTAAAGGATTTCTAGTCT 2238
 Db 2384 GGTGTAAGAAACCTCGAGAGAGGCTGAAGATTTTCTTGGAAATTCAGGATCCAAATAT 2443
 Qy 2239 GAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATATTTGATCTCTTCAAAATAT 2298
 Db 2444 CAAGTCCCAATGCCAAAGAGACTCTCACAGGATTTAAAAAATAATTTACTATTTTGGACCCA 2503

Qy 2299 GGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTGCTGTTGTTAAAAAGGAGTAA 2358
 Db 2504 GGACAACAATACTATTATGGCAGAGCTGAAAAACTATTGCTTTATTAAAGGAGAGTAA 2563

RESULT 8
 US-09-769-787-206
 ; Sequence 206, Application US/09769787
 ; Publication No. US20030091577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Philip M
 ; FILE OF INVENTION: Proteins
 ; TITLE REFERENCE: PWC/P21129WO
 ; CURRENT APPLICATION NUMBER: US/09769,787
 ; PRIOR FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: GB 9816337.1
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: US 60/125164
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 206
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-769-787-206

Query Match 42.0%; Score 1003.8; DB 10; Length 2481;
 Best Local Similarity 66.0%; Pred. No. 6.5e-233;
 Matches 1623; Conservative 0; Mismatches 693; Indels 144; Gaps 6;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTTAAGGAAA---TAATCGTGTTC 57
 Db 60 TTCTTATGAGCTTGGTCTGTCACCAAGCTGGTCAGGATAAGAAAGTCTAATCGAGTTGC 119
 Qy 58 CTATATAGATGGAACAAAGCGACGCAAAAAACGAGAAATTTGACTCCTGTAGAGGTAG 117
 Db 120 TTATATAGATGGTGTATCAGGCTGGTCAAAAGGAGAAACTTGACACAGATGAAGTCAG 179
 Qy 118 CAAGCTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
 Db 180 TAAGAGGGAGGGGATCAACCGCGAAACAAATCGTCATCAAGATTAACGATCAAGGTTATGT 239
 Qy 178 CACTTCACATGCGACCACTATCATTTATACATGTTAGGTTCCCTTATGAGCTATCAT 237
 Db 240 GACCTCTCATGAGAGACCATTAATTTATTAATGCAAGGTCCTTATGATGCCATCAT 299
 Qy 238 CAGTGAAGAAATTTACTCATGAAGATCCAAACCTATAAGCTTAAAGATGAGGATATTGTTAA 297
 Db 300 CAGTGAAGAGCTCCTCATGAAGATCCGAAATTTATCAGTTGAAGATTCAGACATTCGTCNA 359
 Qy 298 TGAGCTCAAGGTGATATGTTATCAAGTAGATGGAATACTACTATGTTTACCTTAAAGGA 357
 Db 360 TGAATCAAGGGTGGTTATGTCATCAAGGTAGACGCAAAATACTATGTTTACCTTAAAGGA 419
 Qy 358 TGCTCCCAACGCGAATAACGTCGTCACAAAGAGGAAATCAATCCGCAAAAACAGAGCA 417
 Db 420 TGCAGCTCATGCGGATAATTTGCGACAAAAGAGAGATTAAACGTCGAGAGCAGGAACG 479
 Qy 418 TAGTCAACATCGTGAAGGTGGAATCCCAAGAAACGATGCTGCTGTTGCTTGGCAGCTTC 477
 Db 480 CAGTCATAATCACGGGTGAGGCT-----AACGATCATGAGTAGCTGAGCCAGAGC 533
 Qy 478 GCAAGGACCTTATCTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGGA 537
 Db 534 CCAAGGACGCTTATCAACGGATGATGGGTATCTTCAATGCATCTGATATCATTTAGGA 593
 Qy 538 TACTGGTATGCTTATATCGTTTCTCATGGAGATCATACCATTCATCTTCAAGATGA 597
 Db 594 CACGGGTGATCTTATATCGTTTCTCAGCGGCGACCATTACCATTTACATTCCTTAAGATGA 653

598 GTTATCAGCTAGGAGTGGCTGCTGCGAGAGCTTCCTATCTGTCGAGGAATCTGTC 657
 654 GTTATCAGCTAGGAGTGGCTGCTGCGAGAGCTTCCTATCTGTCGAGGAATCTGTC 698
 658 AATTCAGAACCTATCGCGACAAATAGCGATAACACATTCAGAACAACTGGGTACC 717
 699 -GAAGCAGGGATCTGCTGCTTCCTTCAAGTCTAGTCTATATAGTAATCCAGCTCAACCA 757
 718 TTCTGTAAAGCAATCCAGAACCTACAAATATAACACAAAGCAACAAACAGCAACACTAACAG 777
 758 GATTCAGAGAACCAATCT-----GACTGTCACTCCAACTTA 797
 778 TCAAGCAAGTCAAAGTAATGACATGTATGCTCTTGAACAGCTCTACAACTGCGCTTT 837
 798 TCATCAAAATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTAAACCCCTT 857
 838 GAGTCAACGACATGTAATCTGATGCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
 858 ATCAAGACGCGATGTGAATCTGATGCGCTTATTTTCGACCCAGCGCAATCACAAGTCG 917
 898 AACAGCTAGAGGTTTTCAGTCCACACGAGATCATTTACCACTTCCTTACTCTCA 957
 918 AACGCCAGAGTGTAGCTGCTCCATGCTGTAACCAATTAACCACTTTATCCCTTATGAACA 977
 958 AATGCTGAATTGAAGAACGAATGCTGCTATTTATTTCCCTTCTGTTATCGTTCAAAACCA 1017
 978 AATGCTGAATTGAAGAACGAATGCTGCTATTTATTTCCCTTCTGTTATCGTTCAAAACCA 1037
 1018 TTGGGTACCAATTCAGGCCAGACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
 1038 TTGGGTACCAATTCAGGCCAGACAAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1097
 1078 AGGCCCGCAACTGCACCAATCTTTAAATAGACTCAA-----ATTCTCTTTT 1125
 1098 AAGTCGCAACTGCACCAATCTTCAACAGCTCCAAAGCAATCCAAATGATGAGAAAT 1157
 1126 GGTATGTCAGTGTACAGAAAGTTGGGAGGATATGTTTCGAGAAAGGCGATCTC 1185
 1158 GGTCAAGAAAGTGTTCGAAAGTAGGCGATGTTATGTTTTCGAGGAGATGGAGTTTC 1217
 1186 TCGTATGCTTTTCGAAAGATTTTACCCTGAAACTGTAAATCTTGAAGCAAGTT 1245
 1218 TCGTATATTCAGCCAGAGATCTTTTACGAGAAACAGACAGAGCAATGATAGCAACT 1277
 1246 ATCAAAACAGAGAGTGTTCACACTTTAACTGCTAAAGAAAGATGTTGCTCTCTCG 1305
 1278 GSCCAAGCAGGAAAGTTTATCTCATAAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG 1337
 1306 TGACCAAGATTTTATGATAAGCATATATCTGTTTAACTGAGGCTCATAAAGCCTTGT 1365
 1338 TGATCGAGAAATTTTCAATTAAGCTTATGACTTATGAGCAAGATTTACCAAGATTTACT 1397
 1366 TGNAAATAGGTCGTAATCTGATTTCCAAAGCTTTAGACAAATTTATAGAACGCTTGA 1425
 1398 TGATTAATAAGTCGACAGATTTGATTTGAGGCTTTGATAACCTGTTGGAACGACTCAA 1457
 1426 TGATGAATCCGATTAATAAGAAATTTGTTAGATGATTTATTTGGCATTTCTAGCAACCAT 1485
 1458 GGATGTCCTCAAGTATAAGTCAAGTTAGTGGATGATTTTTCGCTTCTTAGCTCCGAT 1517
 1486 TACCATTCAGAGCAGCTTTGGCAACCAATTTCTCAATTTAGTATGATTAAGTCAAGCAAGT 1545
 1518 TCGTCAATCAGAGCTTTTAGGAAACCAATGGCAATTTACCTACACTGATGATGAT 1577
 1546 TCGTATGCTCAATTAGCTGATAGTATACAGCTCAGATGTTTACATTTTATGATGAACA 1605
 1578 TCAAGTAGCAAGTTGCGAGGCAAGTACACACAGAGAGCGTTATATCTTTGATCCTCG 1637
 1606 TGATATAATCAGTGAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
 1638 TGATATAACCAAGTGAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697

1666 GATTGAAAGATGAGCTTTCTGATAGGAAAGTTGAGCTCAAGCCTATATCTAAAGA 1725
 1698 GATTAAAGATAGTTTGTCTGAGAGCTGAGAGCGGCGAGCCGCTTATCTAAAGA 1757
 1726 AAAGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCCACTGAGATAG 1785
 1758 GAAAGGTTGACCCCTCTTCGACAGACCAATCAGGATTCAGGAATTAAGTCAAGGCAAGG 1817
 1786 TGACGAGCTTATTTACAATCGTGTGAAAGGGGAAAAACGAAATTCCTCGTTGACCTTCC 1845
 1818 AGCAGAGCTTCTCAACCGCTGAAAGCAGCTAAGAGGTCGCACTTGTATGATGCC 1877
 1846 ATATATGTTTACGATACAGTTGAGGTTTAAAGGTTAAATTTGATTTATTTCTCATAAGA 1905
 1878 TTCAATCTTCAATATATCTGTAAGAGTCAAAACGCTAGTTTAAATCATCTCATTTAGA 1937
 1906 TCATTACCAATTAATAATTTGCTGTTTGTATGATCACACATACAAAGCTCCAAATGG 1965
 1938 CCATTACCAATTAATAATTTGCTGTTTGTATGAGGCTTATGAGGCACTTAAGG 1997
 1966 CTATACCTTGAAGATTTGTTTGCAGCATTAAGTACTACGTAGAAACACCTTGACGACG 2025
 1998 GTATATCTTTGAGGATCTTTTGGGACTGTCAAGTACTATGTCGAACATCCAAACGACG 2057
 2026 TCACATCTTAATGATGATGAGTGGGCAATGCGAGTGCATGTTTAGGCAAGAGACCA 2085
 2058 TCCGATTCAGATAATGTTTGGTAAAGCTAGCGACCATGTTTCAAGAAACAAAAATGG 2117
 2086 CAGTGAAGTCCAAATTAAGACTTCAAGGGGATGA----- 2121
 2118 TCAAGCTGATACCAATCAACCGGAAACCCNAGCGAGGAGAACTCAGACAGAAAAACC 2177
 2122 ----- 2121
 2178 TGAGGAAGAACCCCTCGAGAGAGAAACCGAAACGAGAAACGAGAGTCTCCAAACC 2237
 2122 ---AGAGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAA 2178
 2238 AACAGAGAACCCAGAGAAATCACCAGAGAAATCAGAGAACTCAGAGTCCGAGACTGAA 2297
 2179 AGTAGAGCCCACTCAAGAGAGAGAGAGTGTTCGCGAAAGTAAACGAGTCTTAGTCT 2238
 2298 GGTGAAGAAACCTGAGAGAGCTGAGATTTACTTGGAAATCCAGGATCCCAATTAT 2357
 2239 GAAAGCCCAATGCAACAGAGAACTCTAGCTGTTTACGAAATTAATTGACTCTTCAATAT 2298
 2358 CAGTCCCAATGCCAAGAGACTCTCAGAGATTTAAAAAATAATTTACTATTTTGGCACC 2417
 2299 GATTAACATAGTATCATGCGAGAGAGAGAAATTTCTGCTGTTGTTAAAGAGAGTAA 2358
 2418 GGACAAACATATATTATGGCAGAGCTGAAAAAATTTTGGCTTTTATTAAGAGAGTAA 2477

RESULT 9

US-10-412-862-11
 ; Sequence 11, Application US/10412862
 ; Publication No. US20040052781A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-685
 ; CURRENT APPLICATION NUMBER: US/10/412,862
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 11

```

; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-10-412-862-11

Query Match      41.5%; Score 991; DB 13; Length 2531;
Best Local Similarity 65.7%; Prd. No. 8.5e-230;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAA--TAATCGGTGTTTC 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 TTCCTATGAGCTTGGACGGTTACCAAGCTGGTCAGGATAGAAAAGAGTCTAATCGAGTTGC 119

QY 58 CTATATAGATGAAAACAAAGCGACGCGAAAAACGGAGAAATTGACTCTCTGATGAGGTTAG 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 TTATATAGATGGTGTATCAGGCTGGTCAAAAGGCGAAAAAATTGACACCAGATGAAGTCAG 179

QY 118 CAAAGCGTGAAGGAATCAATGTCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGT 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 TAAAGAGGAGGGGATCAACCGCGAAACAAATTGTTATCAAGATTACGGATCAAGGTTATGT 239

QY 178 CACTTCAATGCGGACCACTATCAATTATTACAATGGTTAAGGTTCTCTATGACGCTATCAT 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 GACCTCTCATGAGACCACTATTCACTTATATAATGGCAAGGTTCTTATGATGCAATCAT 299

QY 238 CAGTGAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAAATGAGGATAATTGTTAA 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CAGTGAAGAGCTTCCTCATGAAGATCCGAATTATCAGTTGAAGGATTGAGACATTGTCAA 359

QY 298 TGAGGTCAAAGGTTGGATGTGTTATCAAGGTAGATGGAATAACTATCTGTTACCTTAAGGA 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TGAATCAAGGTTGGTTATGTCATTAAGTAAACGGTAAATCTATGTTTACCTTAAGGA 419

QY 358 TGCTGCCACCGGATACGTCGTCGTAACAAAGAGAAATCAATCGACAAAAACAAGACCA 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 TCGRCTCATCGGATAATAATTCGGAACAAAGAGAGATTAAACGTCAGAAGCAGGAAACG 479

QY 418 TAGTCAACATCGTCAAGGTGAACTCCAAGAAACGATGGTGCTGTGTTGCCCTTGGCAAGTTC 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 CAGTCATANTCAT-----AACTCAAGAGCAGATAAATGCTGTGCTCGACCCAGAGC 530

QY 478 GCAAGGACGCTATACTACAGATGATGTTATATCTTTAATGCTCTTGATATCATAGGGA 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 CCAAGGACGTTATACAACGGATGATGGTATATCTTCAATGCAATCTGATATCAATTGAGGA 590

QY 538 TACTGTGTATGCTTATATCGTTCCTCATGAGGATCATTAACCATTAATTCCTAAGAAATGA 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 CACGGGTGATGCTTATATCGTTCCTCAACGGCGACCATTAACCATTAATTCCTAAGAAATGA 650

QY 598 GTTATCAGCTAGCGAGTTGGCTGTCGAGAAAGCCTTCTTATCTGGTCGAGGAATCTGTC 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 GTTATCAGCTAGCGAGTTAGCTGCTCGAAGCGCTATTGGAAATGG----- 695

QY 658 AAATTCAGAAACCTATCGCGCAACAAATACGATAACACTTCAAGAAACAAACTGGGTACC 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 696 -GAAGCAGGGATCTCGTCCCTCTTCCAAGTTCTAGTTATTAATGCAATCCAGCTCAACCAA 754

QY 718 TTCTGTAAAGCAATCCAGGACTACAATACTAACAGCAACACAGCAACACTAACAG 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 755 GATTGTGAGAAACCAACATCT-----GACTGTCACTCCAACCTTA 794

QY 778 TCAAGCAAGTCAAAGTAATGACATTGATGTCTCTTGAACAGCTCTCAAACTGCCTTT 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 TCATCAAAATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGTGCTTAACCCCTT 854

QY 838 GAGTCAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 ATCAGAAACGCAATGTGGAAATCTCATGGCCTTATTTTCGACCCAGCGCAATCACAAGTCG 914

QY 898 AACAGCTAGAGGTGTTGCAAGTGCACACGGAGATCATTAACACTTCACTCCCTTACTCTCA 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 AACCCGACAGGTTGAGTGTGCTCATGGTAACCAATTACCACTTTATCCCTTATGAACA 974

```

[illegible]

Db	1329	TTACAAATAAGGCTTATGACTTACTAGCAAGAAATTCCACCAAGATTACTTGTATAATAAAGG	1388
QY	1378	TCGTAATTTCTGATTTCCAAAGCCCTTAGACAAAATTATTAGAACGCTTGATGATGATGCAC	1437
Db	1389	TCGACAAGTTGATTTTGAGGCTTTGGATAACTGTGTGGAAGCACTCAAGGATGTCTCAAG	1448
QY	1438	TAATTAAGAAAATTTGGGTAGATGATTATTATGGCATTCCTAGCACCAATTAACCATCCAGA	1497
Db	1449	TGATAAGTCAAGTTAGTGAAGATATTCTTTGCCCTCTTAGCTCCGATTCGTATCCAGA	1508
QY	1498	GGGACTTGGCAACCAAAATTTCTCAAAATTTAGTATATCTGAAGACGAAGTTTCGTATTTGCTCA	1557
Db	1509	ACGTTTAGAAAACCAAAATGGCAAAATTAACCTACACTGATGAGATTCAAAGTAGCCAA	1568
QY	1558	ATTAGCTGATAGTATACAAAGTCAGATGGTTTACATTTTGTGAGAACATCATATATATCAG	1617
Db	1569	GTTTGCAGCAAGTACACAGCAGAAGACGGTTATATCTTTGATCCTCGTATATAACCCAG	1628
QY	1618	TGATGAAGGAGATGATATGTAACCGCTCATPATGGGCATAGTCACCTGGATTGGAAAAGA	1677
Db	1629	TGATGAGGGGATGCTATGTAATCTCCACATATGACCCATAGCCACTGGATTAAAAAGA	1688
QY	1678	TAGCCTTTCTCATGAAGAAAAGTTGCAGCTCAAGCCCTATATCTTAAGAAAAGGATTCCT	1737
Db	1689	TAGTTTGTCTGAAGCTGAGAGCGGAGCCCAAGGCTTATGCTTAAAGAGAAAGGTTTGAC	1748
QY	1738	ACCTCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACGACTAT	1797
Db	1749	CCCTCCTTCGACAGACCATCAGGATTCAGGAATCTGAGGCGAAAAGAGCAGAACTAT	1808
QY	1798	TTACAATCTGTGTAAGGGGAAAAGAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTCA	1857
Db	1809	CTACAACCGMGTGAAGACAGCTAAGAAGGTGCCACTTGATCGTATGCTTACATCTTCA	1868
QY	1858	GCATACAGTTTGAGCTTAAACCGTAATTTGATTTATTCCTCATAGGATCATTAACATAA	1917
Db	1869	ATATCTGTAGAAGTCAAAAACGGTAGTTTAATCTACCTCATTTATGACCAATTACCATAA	1928
QY	1918	TATTAAATTTGCTTGGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGA	1977
Db	1929	CATCAAAATTTGAGTGGTTTTCGAAAGCCCTTATGAGGCACCTTAGGGGTATACTCTGA	1988
QY	1978	AGATTTGTTTGCAGAGATTAAGTACTACGTAGAACACCTCTGACGACGCTCCACTTCTAA	2037
Db	1989	GGATCTTTTGGCCACTGTCAAGTACTATGTGCAACATCCAAACGAAGCTCCGCAATTCGA	2048
QY	2038	TGATGGATGGGCAATGCGCAGTCAGCATGTGTTAGCCAGAAGACCAAGTGAAGATCC	2097
Db	2049	TATTTGTTTGGTAAACGCTAGCGACCATGTTTCAAGAAACCAAAAATGGTCAAGCTGATAC	2108
QY	2098	AAATAAGAACTTCAAAAGCGATGA	2121
Db	2109	CAATCAAAACGAAAAACCAAGCGAGGAGAAAACCTCAGACAGAAAAACCTGAGGAAGAAAC	2168
QY	2122	-----AGAGCCAGT	2130
Db	2169	CCCTCGAGAAGAGAAACCGCAAGCGAGAAAACGAGTCTCCAAAACCAACAGAGGAACC	2228
QY	2131	AGAGAAAACCTGCTGAGCCAGAGAGTCCCTCAAGTAGAGACTGAAAAGCTAGAGGCCCA	2190
Db	2229	AGAAGAAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAGACTGAAAAGGTTGAAGAAA	2288
QY	2191	ACTCAAGAGACGAAGTTTTCCTTTCGAAAAGTAACGGATTCAGTCTGAAAAGCCAAATGC	2250
Db	2289	ACTGAGAGAGGCTGAAGATTTACTTTGGAAAAATCCAGGATCCAATTCAAGTCCAATGC	2348
QY	2251	ACAGAAAACCTAGCTGTTTACGAAAATAATTTGACTCTTCAAAATTATGGATAACAATAG	2310
Db	2349	CAAAAGAGCTCTCACAGGATTAATAAAATAATTTACTATTTCGCACCCAGGACACAATAC	2408
QY	2311	TATCATGGCAGAGCAGAAAAATTAATTCGCTGTGTTTAAAGAGATGA	2358

Db 2409 TTATTGGCAGAGCTGAAAACTATTGGCTTTTATTAAAGGAGATAA 2456

RESULT 11
US-10-387-783-11
; Sequence 11, Application US/10387783
; Publication NO. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-11

Query Match 41.5%; Score 991; DB 16; Length 2531;
Best Local Similarity 65.7%; Pred. No. 8.5e-230;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTTC	57
Db	60	TTCTATGAGCTTGGAGCTTACCAAGCTGGTCAGATAAGAAAGAGTCTAATCGAGTTGC	119
Qy	58	CTATATAGATGGAAAAACAAGCGACGCAAAAAACGGAGAAATTGACTCCTGATGAGGTTAG	117
Db	120	TTATATAGATGGTGTATCAGGCTGGTCAAAAGGCGAANAACCTTGACACGAGATGAAGTCAG	179
Qy	118	CAAGCGTGAAGGATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT	177
Db	180	TAAGAGGAGGGGATCAACGCCAACAAATGGTTATCAAGATTACGGATCAAGGTTATGT	239
Qy	178	CACCTTCACATGGCGACCACTATCATTTATTACAATGGTAAGGTTCTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGACCATTTATCATTTACTATATATGCGAAGGTTCTTATGATGCCATCAT	299
Qy	238	CAGTGAAGAATTTACTCATGAAGATCCAACTATAAGCTAAAAAGTAGGAGATATTGTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTTATAGTTGAAGATTCACACATTGTCAA	359
Qy	298	TGAGGTCAAGGGTGGATATGTTATCAAGTAGATGGAATACTATGTTTACCTTAAGGA	357
Db	360	TGAATCAAGGGTGGTTATGTCAATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA	419
Qy	358	TGTCGCCACGGCGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAAAACAAGCA	417
Db	420	TGCRGCTCATGCGGATAATTTCCGCAAAAGAGAGATTAACGTCAGAGCAGGAACG	479
Qy	418	TAGTCAACATCGTAAGGTGGAACTCTCAAGAAACGATGGTGTGTTGCCCTTGGCAGCTTC	477
Db	480	CAGTCAATATCAT-----AACTCAAGAGCAGATAATGCTGTGTGTCGACCCAGAC	530
Qy	478	GCAAGGACGCTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGA	537
Db	531	CCAAGACGTTATAACCGGATGATGGGTATATCTTCAATGCATCTGATATCATTTGAGA	590
Qy	538	TACTGGTGATGCTTATATCGTTCCCTCATGGAGATCAATACCATTACATTCCTTAAGATCA	597
Db	591	CACGGGTGATGCTTATATCGTTCTTCAACGGCGACCATTTACCAATTACATTCCTTAAGATCA	650
Qy	598	GTTATACGCTACGAGTTGGCTGCTGCAGAAACCTTCCTATCTGTCCGAGGAAATCTGTTC	657

```

1689 TAGTTTGTCTGAAGCTGAGAGCGCGAGCCAGCCAGGCTTATCTRAAGAGAAAGTTTGAC 1748
1738 ACCTCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACAGCTAT 1797
1749 CCCTCTCTCGACAGACCAATCAGGATTCAGGAAATCTAGGAGCAAAAGGAGCAGAGCTAT 1808
1798 TTAACAATCTGTGTGAAGGGGAAAGAAAGCAATTCACCTCGTTCCGACTTCCATATATGGTTGA 1857
1809 CTACAACCGGTGTAAGAGCGCTAAGAGGTCCTGATGATGCTTGAATGCTTCAATCTTCA 1868
1858 GCATACAGTTGAGTTAAAGCGTAAATTTGATTTATTCCTCATAAAGGATCATTAACCAATA 1917
1869 ATATACCTGTGAAGAGTCAAAACCGTAGTTTAAATCAATACCTCATTAATGACCAATTA 1928
1918 TATTAAATTTGCTTGGTTGATGATCACAATCAAAAGCTCCAAATGGCTATATACCTTGA 1977
1929 CATCAATTTGAGTTGTTGACGAAGGCTTTATAGGCGACCTTAAGGGGTATATCTTTGA 1988
1978 AGATTTGTTGCGACGATTAAAGTACTAGTACGAGACCTTGAAGAGCTCCACATTTCTAA 2037
1989 GGATCTTTTGGCGACTGTCAGTACTATGTCGAACATCCAAACGAAAGTCCGCAATTCAGA 2048
2038 TGATGATGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAAGCACCACAGTGAAGATCC 2097
2049 TAATGGTTTGGTAACTGCTAGCGACCATGTTCAAGAAACAAATAATGCTCAAGCTGATAC 2108
2098 AAATAAGAACTTCAAAGCGGATGA----- 2121
2109 CAATCAAGCGAAAGAAACCAAGCGAGGAGAAACCTCAGACAGAAAAAOCCTGAGGAGAAAC 2168
2122 -----AGAGCCAGT 2130
2169 CCCTCGAAGAGAGAAACCGCAAGCGAGAAACAGAGATCTCCAAACCAACAGAGAAC 2228
2131 AGAGAAACACCTGCTGAGCCAGAGAGTCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCA 2190
2229 AGAAGATCACAGAGGAAATCAGAGAACCTCAGTGCAGACTGAAAAAGTTGAAGAAA 2288
2191 ACTCAAGAAAGCAGAGTTTGTCTGCGAAAGTAAGGATTCAGTCTGAAAGCAAGTCC 2250
2289 ACTGAGAGAGCTGAAGATTTACTTTGAAAAATCCAGATCCAAATATCAAGTCCATGC 2348
2251 AACAGAACTCTAGCTGCTTACGAAATAATTTGACTCTTCAATATATGGAATAACAATAG 2310
2349 CAAGAGAGCTCTCAGAGATTAATAAATAATTTACTATTTGCAACCCAGGACACAATAC 2408
2311 TATCATGGCAGAGCAGAGAAATTTACTTTGCGTTGTTAAAGGAAGTAA 2358
2409 TATTATGGCAGAGCTGAAATACTATTGCTTTTAAAGGAGAGTAA 2456

RESULT 12
US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA

```


; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

Query Match	41.4%	Score 990;	DB 13;	Length 2531;
Best Local Similarity	68.1%	Pred. No. 1.5e-229;		
Matches 1478; Conservative	0;	Mismatches 636;	Indels 57	

QY	1	TTCTTACGAGTTGGACCTGTATCAAGCTAGAACGTTAAGGAAA---TAATCGTGTTC	57
Db	60	TTCCATATGAACCTTGGTCGTACCAACAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC	119
QY	58	CTATATAGATTGGAATAACAAGCGACGCAAAACCGAGNAATTGGACTCTGTATGAGGTTAG	117
Db	120	TTATATAGATTGGTCATCAGGCTGGTCAAAAGCGCAGAAACTTGCACACAGATGAAGTCAG	179
QY	118	CAAGCGTGAAGGAATCAATCTGTAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGT	177
Db	180	TAAGAGGAGGAGGATCAACGCGGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGT	239
QY	178	CACCTTCATGGCGACCACTATCATTTATACAATGGTAAAGTTCCTTATGACGCTATCAT	237
Db	240	GACCTCTCATGGAGACCATTTATCATTTACTATTAATGGCAAGTCCCTTATGATGCCATCAT	299
QY	238	CAGTGAAGAATTACTCATGAAGAATCCAAACTATAAGCTAAAGATCAGGATATTGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAATTATCAGTTGAAGATTTCAGACATTGTCAA	359
QY	298	TGAGGTCAAGGTTGGATATGTTTATCAAGCTAGATGGAAATATCTATGTTTACCTTAAGGA	357
Db	360	TGAATTCAGGGTGGTTATGTTTATCAAGGTAGATGGAAATATCTATGTTTACCTTAAGGA	419
QY	358	TGCTGCCACGCGGATAACGTCGCTCAAAAGAGGAAATCAATCGACAAAAACAAGACGA	417
Db	420	TGCAGCTCATGCGGATAATATTCGGACAAAAGAGAGATTAAAGTCAGAAAGCAGGAACA	479
QY	418	TAGTCAACATCTGAAGGTGGAACTCAAGAAACGATGGTGTGTGCTTGCCTTGCAAGTTC	477
Db	480	CAGTCAATATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCAGCCAGAGC	533
QY	478	GCAAGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGA	537
Db	534	CCAAGGACGCTATACAACGGATGATGGTTATATCTTCAATGATCTGATATCATTTAGGGA	593
QY	538	TACTGTTGATGTTATATCGTTTCTCATGGAGATCAATTACCATTCATTCCTTAAGAATGA	597
Db	594	CACGGGTGATGTTATATCGTTCTCCTCAGCGGACCAATTACCATTCATTCCTTAAGAATGA	653
QY	598	GTTATCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCTTCTATCTGTGCGAGGAATCTGTC	657
Db	654	GTTATCAGCTAGCGAGTTAGCTGCTCAGAGGCT-----	688
QY	658	AAATTCAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACC	717
Db	689	-----ATTGGAATGGGAACGAGGATCTGTCTCTTCAAAGTTCTAGTTTATAATGC	740
QY	718	TTCTGTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACAGCAGCAACACTAACAG	777
Db	741	AAATCCAGCTCAACCAAGATTGTCAGAGAACCAAACTGACTGTCACTTCCAACCTTATCA	800
QY	778	TCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTTGAAACAGCTCTACAAACGTGCTTT	837
Db	801	TCA---AAATCAAGGGGAAAAATTTCAAGCCTTTTACGTGAATGCTATGCTTAACCCCTT	857
QY	838	GAGTCAACGACATGTAGAAATCTGATGGCCCTGCTTTGATCAGACCAAAATCAAGTCG	897
Db	858	ATCAGAACCCATGTGGAAATCTGATGGCCCTTATTTTCGACCCAGCGCAATCACAGTCG	917
QY	898	AACAGCTAGAGGTGTGCAAGTCCCAACAGGAGATCAATTACCACTTTCATCCCTTACTCTCA	957
Db	918	AACGCCAGAGGTGTAGCTGTCCCTCATGTGTAACCATTACCACTTATCCCTTATGAACA	977
QY	958	AATGTTCAATTTGGAAGAACGNAATCGTCGTATATATTCCTCTCGTTATPCTGTTCAAACA	1017

QY	2086	CAGTGAAGATCCAAATAAGAACTTCAAAGCGGATGAAGCCAGTAGTAGGAAACACCTGC	2145
Db	2118	CCAAGACAGTAAACCTGATGAAGATAGGAACATGATGAAGTAGTGAGCCAACTCACCC	2177
QY	2146	TGAGCCAGAAG	2156
Db	2178	TGAATCTGATG	2188

RESULT 13
US-10-412-850-5
; Sequence 5, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Motifins
; TITLE OF INVENTION: Motifins
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-5

QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGNAAA---TAATCGTGTTC	57
Db	60	TTCTTATGAACCTTGGTCTGCTCACCAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTC	119
QY	58	CTATATAGATGAAACACAGCGACGCAAAAACGAGAAATTTGACTCTCTGATGAGGTTAG	117
Db	120	TTATATAGATGGTATCAGGCTGGTCAAAAGCAGAAACTTGACACCAAGATGAAGTCAG	179
QY	118	CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATPACAGACCAAGGCTATGT	177
Db	180	TAAGAGGAGGGGATCAACGCCGAAACAAATCGTCATCAAGATTACGGATCAAGTTATGT	239
QY	178	CATTCAATGCGGACCATCATTTTACAATGGTAGGTTCCCTTATGACGCTATCAT	237
Db	240	GACCTCTCATGAGACCATTTACTTACTTATATGSCAAGGTCCTTTATGATGCCATCAT	299
QY	238	CAGTGAAGATTTACTCATGAAGATCCAACTATAAGCTATAAAGATGAGGATTTGTTAA	297
Db	300	CAGTGAAGAGCTCCTCATGAAGATCCGAAATTTATCAGTTGAAGATTACAGATTTGTC	359
QY	298	TGAGGTTCAAGGGTGGATGTTTATCAAGGTAGATGGAATAATCTATGTTTACTTTAAGGA	357
Db	360	TGAAATCAAGGGTGGTTTATGTTTATCAAGGTAGATGGAATAATCTATGTTTACTTTAAGGA	419
QY	358	TGCTGCCACGCGGATAACGCTCGTACAAAGAGGAATCAATCGACAAAACACAGGACA	417
Db	420	TGCAGCTCATGCGGATAATATTCGGAACAAAGAGAGATTTAAACGTCAGACGAGGAA	479
QY	418	TAGTCAACATCGTGAAGGTGGAATCCAGAAAACGATGGTGTCTGTTCCCTTTGACGTTTC	477
Db	480	CAGTCATATCACGGGGTGGTCT-----ACGATCAAGCAGTAGTTGACGCCAGAGC	533
QY	478	GCAAGGACGCTATACTACAGATGATGGTTATCTTTTAATGCTTCTTGATATCATAGGA	537

Query Match 41.4%; Score 990; DB 16; Length 2531;
Best Local Similarity 68.1%; Pred. No. 1.5e-229;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

1126 GGTAGTACGCTGGTAAAGCTGGGGAAGGATATGATTCGAGAAAGGCACTTC 1185
 1158 GGTAAAGAGCTGTTGAAAGAGTAGGCGATGTTATGTTCTTTGAGGAGATGGAGTTTC 1217
 1186 TCGTTATGCTTTGCGAAGATTTACCATCTGAACCTGTTAAATCTTGAACCAAGTT 1245
 1218 TCGTTATATCCAGCCAGGATCTTTGAGGAGAAACAGCAGCGCATGATGACAACT 1277
 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCG 1305
 1278 GGCAAGAGAGAGTGTTCATCTAAGCTAGGAGCTTAAGAAAGTACCTCCCATCTAG 1337
 1306 TGACCAAGAGATTTATGATAAGCATATAATCTGTTAACTGAGGCTTAAGAGCTTTT 1365
 1338 TGATCGAGATTTTACAAAGGCTTATGACTTACTAGCAAGATTTACCAAGATTTACT 1397
 1366 TGNAAATAGGCTGTTATCTGATTTCCAGGCTTTAGACAAATTTATGAAAGCTTCAA 1425
 1398 TGATAATAAGAGTGCACAAAGTTGATTTTGGGCTTTGGATTAACCTGTTGGAGCACTCAA 1457
 1426 TGATCAATCGACTAATAAGAAAGATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAAT 1485
 1458 GGATGTCCTCAAGTGAATAAGTCAAGTGTAGTGAATATTTCTGCTTTCTAGCTCCGAT 1517
 1486 TACCATCCAGAGGCTTGGCAACCAAAATTTCTCAATTTGAGTATCTAGAGCAAGT 1545
 1518 TCGTCATCCAGAACGTTTGGGAAACCAAAATTCGCAAAATTTACTACCTGATGATGAT 1577
 1546 TCGTATGCTCAATTTAGCTGATTAAGTATACAGCTCAGATGTTTATTTGATGAACA 1605
 1578 TCAAGTAGCAAGTTGGCAGGCAAGTACACAGAGAGGTTTATCTTTGATCTCTCG 1637
 1606 TGATATAATCACTGATGAAGAGATGTCATATGATTAAGCTTATATGGCCATATGACCTG 1665
 1638 TGATATAACCACTGATGAGGAGGATGCTTATGTAATCTCCATATGACCCATAGCCACTG 1697
 1666 GATTGGAAGATGAGCTTTCTGATTAAGAAAGTTGAGCTCAAGCTTATCTAAAGCA 1725
 1698 GATTAAAGAAAGTAGTTTCTGTAAGCTGAGAGCGGCGAGGCTTATGCTAAAGCA 1757
 1726 AAAAGTATCCTACTCCATCTCCAGACGAGATGTTAAAGCAAACTTCAACTGGAGATAG 1785
 1758 GAAAGCTTTGACCCCTCTTCCAGACAGCAATCAGGATCAGGAAATCTGAGCAAGAG 1817
 1786 TGACGAGCTATTTTCAATCTGTGAAAGGGGAAAGCAATTTCCACTCGTTCGACTTCC 1845
 1818 AGCAGAGCTATCTCAACCGCTGAAAGCAGCTAAGAGGTCGCACTTGATCGTATGCC 1877
 1846 ATATATGTTGAGCATACAGTTGAGTTTAAAGCGTAAATTTGATTTATCTCTCATAGGA 1905
 1878 TTACAATCTTCAATATCTAGAGGTCGCAAAAGCGGTAGTTTATCATCTCATATGA 1937
 1906 TCAATACCAATATTAATTTGCTGTTGATGATCACATACAAAGCTCCAAATGG 1965
 1938 CCATACCAATCAATCAATTTGAGTGTGTTGAGAGGCTTTATGAGGCACTTAAGGG 1997
 1966 CTATACCTTGGAGATTTGTTGCGACGATTAAGTACTGAGTGAACACCTGACGAGC 2025
 1998 GTATCTCTTGGAGATCTTTTGGCGAGTGTCAAGTACTATGTGCAATCCAAACGAGC 2057
 2026 TCCATCTTAAATGATGAGTGGGCAATGCCAGTGAGCAGTGTGTTAGGCAAGAGACCA 2085
 2058 TCCGATTCAGATAATGTTTGTGTAAGCTAGGAGCCATGTTCTGTAATAATAGGTAGA 2117
 2086 CAGTGAAGATCCAAATAGAACTTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGC 2145
 2118 CCAAGACAGTAACTGATGAAGATAGGAACATGATGAAGTGAAGTGAAGTGAAGTGAAG 2177
 2146 TGAGCCAGAG 2156
 2178 TGAATCTGATG 2188

RESULT 15
 US-09-765-272-65
 ; Sequence 65, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 65:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2290 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
 ; US-09-765-272-65
 Query Match 41.3%; Score 987.6; DB 9; Length 2290;
 Best Local Similarity 67.7%; Pred. No. 5.3e-229;
 Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
 QY 1 TTCTTACGAGTTGGGCTGCTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
 DB 4 TTCTTACGAGTTGGGCTGCTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 63
 QY 58 CTATATAGATGGAAGAAACAAAGCGACGCAAAAGAAACGAGAAATTTGACTCTCTGATGAGTTAG 117
 DB 64 TTATATAGATGTTGATCAGCTGCTGTTCAAAAGGCGAGAAACCTTGACACCAAGATGAAGTCAG 123
 QY 118 CAAGCGCTGAGGAAATCAATGCTGAGCAAAATCGTTCATCAAGATTAACAGATTAACGATTAAGT 177
 DB 124 TAAGAGGAGGAGGATCAACCGCAAAATGTTATCAAGATTAACGATTAACGATTAAGT 183
 QY 178 CACTTCACTGAGGAGGAGGATCAATGCTGAGCAAAATCGTTCATCAAGATTAACGATTAACGATTAAGT 237
 DB 184 GACCTCTCATGAGGAGGAGGATCAATGCTGAGCAAAATCGTTCATCAAGATTAACGATTAACGATTAAGT 243
 QY 238 CAGTGAAGAAATTTACTCATGAAAGATCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGT 297
 DB 244 CAGTGAAGAGCTCTCTCATGAAAGATCGAAATTTACTGTTCAAGGATTAACGATTAACGATTAAGT 303
 QY 298 TGAGGTCAAGGCTGAGATGTTTATCAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 357
 DB 304 TGAATCAAGGCTGTTGTTATGTTCAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 363
 QY 358 TGCTGCCCGCGGATTAACGCTCGCTGACAAAGAGGAAATCAATCGACAAAGAGCA 417

Db 364 TGCAGCTCATGGGATAATATTCGGACAAAGAGATTAACGCTCAGAGCGAAGC 423
QY 418 TAGTCAACATCGTGAAGCTGGACCTCCAGAAACGATGGTGTCTGTTCCTGGCAGCTC 477
Db 424 CAGTCATATATCAT-----AACTCAAGACGACGATTAATGCTGTGTGCGCAGCAGC 474
QY 478 GCAAGGACGCTATATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGA 537
Db 475 CCAAGGACGTTATATCAACGATGATGGGTATATCTTCAATGCACTGATATCATTTGAGGA 534
QY 538 TACTGGTATGCTTATATCTGTTCTCATGAGATCAATTAACATTAATTCCTAAGAAATGA 597
Db 535 CACGGGTATGCTTATATCTGTTCTCATGCGGACCAATTAACATTAATTCCTAAGAAATGA 594
QY 598 GTTATACAGTACGAGTGGCTGTGTCAGAGACCTTCTATCTCGTTCGAGGAAATCTGTC 657
Db 595 GTTATACAGTACGAGTGGCTGTGTCAGAGACCTT----- 629
QY 658 AAATCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACC 717
Db 630 -----ATTGGAATGGAGCAGGAGTCTCGTCTCTTCTCAAGTTCCTAGTTAATATGC 681
QY 718 TTCTGTAAGCAATCCAGGACCTACAATTAACACAAAGCAACCAACGACCACTAACAG 777
Db 682 AAATCCAGCTCAACCAAGATTTGTGAGAGAACCAATCTGACTCTCACTCCAACTTATCA 741
QY 778 TCAAGCAAGTCAAAAGTAAATGACATTTGATGCTCTTGAACAGCTCTACAAACTGCCCTT 837
Db 742 TCA---AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCT 798
QY 838 GAGTCAACGATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAGTCG 897
Db 799 ATCAGAACGCCATGTGGAATCTGATGSCCTTATTTGACCCAGCGCAATCAAGTCG 858
QY 898 AACAGCTAGAGTGTTCAGTGGCAACAGGAGATCAATTAACATTCCTTACTCTCA 957
Db 859 AACCGCCAGAGTGTAGCTGTCTGCTCATGTGAACCAATTAACCTTTATCCCTTATGAACA 918
QY 958 AATGTCTGAATTTGGAAGAACGAATCGCTGTATTTATTTCCCTTCTGTTATCGTTCAACCA 1017
Db 919 AATGTCTGAATTTGGAAGAACGAATCGCTGTATTTATTTCCCTTCTGTTATCGTTCAACCA 978
QY 1018 TTGGGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
Db 979 TTGGGTACAGATTCAGAACCAAGCAACCAAGTCCACCAATCGACTCCGGAACCTAGTCC 1038
QY 1078 AGSCCGCACTGCAACCAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
Db 1039 AAGTCGCAACCTGCAACCAATCTTCAACGAGCTCCAGCAATCCAATTTGATGAGAAAT 1098
QY 1126 GGTATGTCAGCTGGTACGAAAGTTCGGGAAGGATATGTTTCGAGAAAGGGCATCTC 1185
Db 1099 GGTCAAGAGCTGTTGGAAGAGTAGGCGATGTTATGCTTTGAGGAGAAATGGAGTTTC 1158
QY 1186 TCGTTATGCTTTGGGAAAGATTTACCAATCTGAACTGTAAATAATCTTGAAGCAAGTT 1245
Db 1159 TCGTTATATCCAGCAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAACT 1218
QY 1246 ATCAAAACAGAGAGTGTTCACACCTTAATCTGCTAAAAAGAAATTTGTCCTCG 1305
Db 1219 GGCCAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG 1278
QY 1306 TGACCAAGAAATTTATGATAAGCATATAATCTGTTAATCTGAGGCTCATAAAGCTTGT 1365
Db 1279 TGATCGAGATTTTCAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 1338
QY 1366 TGNAAATAGGTCGTAATTTCTGATTTCCAGCTTTAGACAAATATTAGAACGCTTGA 1425
Db 1339 TGATAATAAGGTCGACAGGTTGATTTTTCAGGCTTTGGATACCTGTTGGAAAGCACTCAA 1398
QY 1426 TGATCAATCGACTAATAAGAAATTTGGTAGATCAATTTATTTGGCAATCTCTAGCAACCAAT 1485

Db 1399 GGATGTCNCAAGTGAATAAAGTCAAGTTAGTGGANGATATTTCTTGCTTTCTTAGCTCCGAT 1458
QY 1486 TACCCATCCAGAGGACTTGGCAACCAAAATTTCTCAAAATTTAGTATATCTGAGAGCAAGT 1545
Db 1459 TCGTCATCCAGAACGTTTAGGAAACCAATTCGCGAAATTTACCTACACTGATGATGAGAT 1518
QY 1546 TCGTATGCTCAATTAGCTGATTAAGTATACAAAGTTCAGATGGTTTACATTTTGTATGAACA 1605
Db 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAAGAACAGCGTTATATCTTTGATCTCTG 1578
QY 1606 TGATATATCAGTATGATGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTG 1665
Db 1579 TGATATACCAAGTATGATGAGGGGATGCCATGTAACCTCCATATGACCCCATAGCCTG 1638
QY 1666 GATTGGAAAAGATAGCCCTTCTGATAAGGAAAAAGTTTGACGCTCAAGCTTACTATAAGA 1725
Db 1639 GATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGCAGCCAGGCTTATGCTAAAGA 1698
QY 1726 AAAAGGTATCTACCTCCATCTCCAGACGCGAGATGTTTAAAGCAAAATCCAACTGGAGATG 1785
Db 1699 GAAAGGTTTGACCCCTCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAG 1758
QY 1786 TGCAGCAGCTATTTACCAATCGTGTGAAGGGGAAAAAGAAATTCCTACTGTTTGCATCTCC 1845
Db 1759 AGCAGAAGCTATCTACAAACCGGTGAAGCAGCTTAAGAGGTGCCACTTGTATGCTATGCC 1818
QY 1846 ATATATGTTTGAAGCATAAGTTCAGGTTTAAAGCGGTAAATTTGATTTATCTCTCATAGGA 1905
Db 1819 TTCAATCTTCAATATACTCTAGAGTCAAAACGGTGTAGTTTATCATCTCATTTATGA 1878
QY 1906 TCATTACCAATAATTAATTTGCTTGGTTTGTATGATCAACATACAAAGCTCCAAATGG 1965
Db 1879 CCATTTACCAATAACATCAAAATTTGAGTGGTTTGAAGGCGCTTTATGAGCACCTAAGGG 1938
QY 1966 CTATACCTTGAAGATTTGTTTGGGACGATTAAGTACTAGTGAACACCTGACGAAAG 2025
Db 1939 GTATCTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAAG 1998
QY 2026 TCACATCTTAATGATGATGGGCAATCCAGTGAAGCATGTTAGGCAAGAAAGACCA 2085
Db 1999 TCGCATTCAGATAATAGTTTGGTAAACGCTAGCAGCATGTTCAAGAAACAAAAATGG 2058
QY 2086 CAGTGAAGATCCAAATAAGAACTTCAAGGGGATGAAGCCAGTGAAGAAACACCTGC 2145
Db 2059 TCAAGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAAACC 2118
QY 2146 TGAGCCAGAAAGTCCCTCAAGTAGAGA 2171
Db 2119 TGAGGAGAAACCCCTCGAGAAAGAGA 2144

RESULT 16

US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA

Fri Oct 1 15:35:47 2004

ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Query Match 41.1%; Score 980.8; DB 10; Length 2639;
Best Local Similarity 67.8%; Pred. No. 2.6e-227;
Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;

QY	1	TTCTTACGAGTTGGAGCTGATACAGCTAGAACGGTTAAGGAAA---TAATCGTGTTC	57
DB	173	TTCTATGAACTTGGTGGTCAACCAAGCTGGTCAAGTTAAGAAAGAGTCTAATCGAGTTC	232
QY	58	CTATATAGATGGAACCAAGCGACGCAAAAACGGAGAAATTGACTCCTGATGAGGTAG	117
DB	233	TTATATAGATGGTATCAGGCTGGTCAAAAGCGAGAAATTTGACACAGATGAATCAG	292
QY	118	CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGCTATGT	177
DB	293	TAAGAGAGGGATCAACGCGCAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGT	352
QY	178	CATTTCATGCGGACCACTATCATTTAATGTTAAGGTTCCCTTATGACGCTATCAT	237
DB	353	GACCTCTCATGAGACCAATATCATTTACTAATAGGCAAGGTTCTTATGATGCCATCAT	412
QY	238	CAGTGAAGAAATTACTCATGAAGATCCAAAATATAAGCTAAAGATGAGGATATGTTAA	297
DB	413	CAGTGAAGAACTTCTCATGAAGATCCGAATTAATCAAGTTGAAGGATTACAGACATGTCAA	472
QY	298	TCAGGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGA	357
DB	473	TGAAATCAAGGTTGCTATGTATTAAGTAGACGGAATAACTATGTTTACCTTAAAGA	532
QY	358	TGCTGCCACGGGTAACGTCCTGACAAAGAGGAATCAATCGACAAAACAAAGAGCA	417
DB	533	TCGCGCCCATCGGCAATATTCGACAAAGAGAGATTAACGTCAAGACGAGAACCA	592
QY	418	TAGTCAACATCTGTAAGGTGGAATCCAAAGAACGATGGTGTGTTGCTTGGCACGTTTC	477
DB	593	CAGTCATATCAT------AACTCAAGACAGATAATGCTGTGTCGACGACAGC	643
QY	478	GCAAGACGCTATCTACAGATGATGGTTATATCTTTATGTTCTGATATCATAGAGA	537
DB	644	CCAAGACGTTTATCAACGGATGATGGTATATCTTCAATGATCTGATATCATTTAGGA	703
QY	538	TAATGTTGATGTTATATGTTCTTCATGAGATCATTAACCATTAATCTTAAGAAATGA	597
DB	704	CACGGTGATGTTATGTTCTTCACGGGACCATTAACCATTAATCTTAAGAAATGA	763
QY	598	GTTATCAGCTAGCGAGTTGGTGTGCTGCAAGACGCTTCTTATCTGCTCGAGGAAATCTGTC	657
DB	764	GTTATCAGCTAGCGAGTTAGCTGCTGCAAGGCT-----	798
QY	658	AAATTCAGAACCTATCGCCGCAAAATAGCGAATACATTCAGAACAACTGGGTACC	717
DB	799	-----ATTGGAATGGGAAGCAGGGATCTGCTCCTCTTCAAGTTCTAGTTATATATGC	850
QY	718	TTCTGTAGCAATCCAGGAATACAAATACAAAGCAACAAAGCAACAACTAACAG	777
DB	851	AAATCCAGTTCAACCAAGATTTGTCAGAGAACCAATCTGACTGTCACTCCCACTTATCA	910
QY	778	TCAAGCAAGTCAAGTAATGACATTTGATGATGCTCTTTGAAACAGCTCTACAACTGCCTTT	837
DB	911	TCA--AAATCAAGGGGAAAACATTTCAAGCCCTTTTACGTTGATGCTTAAAGCCCTT	967
QY	838	GAGTCAACGATGATGATCTGATGCGCTTGTCTTTGATCCAGGCAAAATCAAGTCG	897
DB	968	ATCAGAACGCGATGATGATCTGATGCGCTTATTTTCGACCCGCGCAAAATCAAGTCG	1027
QY	898	AACAGCTAGAGGTTGTCAGTGGCCACAGGAGATCAATTAACCTTCTATCCCTTACTCTCA	957
DB	1028	AACGCGCAGAGTTGATGCTGCTCTCATGTTAACCATTACCACTTTATCCCTTATGAACA	1087
QY	958	AATGCTGAAATGGAAGAACGAATCGCTCGTATATATTCCTGCTGTTATCGTTTCAACCA	1017

DB	1088	AATGCTGTAATTGGAACCAAGAAATTCGTGATATATTCCTTCCTGTTATCGTTCAACCA	1147
QY	1018	TTGGGTACAGATTCAAGGCCAGAACCAACCAAGTCCACAAACCACTCCGGAACCTAGTCC	1077
DB	1148	TTGGGTACAGATTCAAGGCCAGAACCAACCAAGTCCACAACTGACTCCGGAACCTAGTCC	1207
QY	1078	AGCCCGCAACCTGCGACCAAAATCTTAAATAGACTCAAAATCTTC-----TTT	1125
DB	1208	AAGTCTCAACCTGCGACCAAAATCTTCAACCAAGTCCAAAGCAATCCAAATGATGAGAAAT	1267
QY	1126	GGTTAGTCACTGGTACGAAAGTTGGGAAGGATATGTTATCGAAGAAAGGCACTCTC	1185
DB	1268	GGTCAAAAGAGCTGTTGCAAAAGTAGCGGATGTTATGTTCTTTGAGAGAAATGAGTTTC	1327
QY	1186	TCGTTATGTTCTTTCGCAAAAGATTTTACCATCTGAAACTGTTTAAAAATCTTGAACCAAGTT	1245
DB	1328	TCGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGCAGGATTTATGATCAAACT	1387
QY	1246	ATCAAAAACAGAGAGTGTTCACACACTTTTAACTTAACTGCTTAACTGAGGCTCATAAAGCTTGT	1305
DB	1388	GGCCCAAGCAGGAAAGTTTATCTCATTAAGCTAGGAGCTAAGAAACTGACCTCCCATCTAG	1447
QY	1306	TGACCAAGAAATTTTATGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGT	1365
DB	1448	TGATCGAAGATTTTACATTAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACT	1507
QY	1366	TGNAATAAGGCTCGTAAATCTGATTTTCCAAAGCTTTAGACAAAATTTATAGAACGTTGAA	1425
DB	1508	TGATAATAAAGTTCGACAAAGTTGATTTTGAAGTTTGGATTAACCTTTGGAACTGCA	1567
QY	1426	TGATGATCGACTAATAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCACCAAT	1485
DB	1568	GGATGCTTCAAGTGAATAAGTCAAGTTAGTGGATGATATTTCTTGCTTTCTTAGCTCCGAT	1627
QY	1486	TACCATCCAGAGGACTTGGCAACCAATTTCTCAAAATTCGATGATGATGATGATGATGAT	1545
DB	1628	TCGTATCCAGAACGTTTAGAANAACCAATTCGCGCAATTAACCTACACTGATGATGAT	1687
QY	1546	TCGTTATGCTCAATTTAGCTGATTAAGTATACAACTGATGATGATGATGATGATGATGAT	1605
DB	1688	TCAAGTAGCAAGTTGGCAGGCAAGTACACAAACAGAGACGGTTATATCTTTGATCCTCG	1747
QY	1606	TGATATATCAGTATGAGGAGATGATATGTTAACTGATGATGATGATGATGATGATGATGAT	1665
DB	1748	TGATATAACCAAGTATGAGGGGATGCTTATGTAATCTCAATATGACCCATAGCCACTG	1807
QY	1666	GATTCGAAAAGATAGCTTTCTGATAAGGAAAGTTGAGCTCAAGCTATATCTAAAGA	1725
DB	1808	GATTAATAAAGATAGTTTCTGTAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTAAAGA	1867
QY	1726	AAAAGGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAG	1785
DB	1868	GAAAGTTTGGACCTCTCTTCGACAGACCCAGGATTCAGGAAATCTAGGCAAAAGG	1927
QY	1786	TGACGAGCTATTTACAACTCGTGAAGGGGAAACCAATTCACCTGCTTCGACTTC	1845
DB	1928	AGCAGAGAGTATCTACACCGGTGAAGAGCAGTAAAGAGTGCCTTGTATGATGCT	1987
QY	1846	ATATATGTTGAGCATACAGTTGAGGTTAAACCGGTAATTTGATTTATTTCTCATAGGA	1905
DB	1988	TTACATCTTCAATATCTGTAAGAGTCAAAACCGGTAGTTTAACTCATCTCATTTATGA	2047
QY	1906	TCATTAACCATATATTTAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGG	1965
DB	2048	CCATTAACCAATCAATCAAAATTTGAGTGGTTTTCGAGGAGGCTTTTATGAGGACCTTAGGG	2107
QY	1966	CTATACCTTTGGAGATTTGTTTCGACAGATTAAGTACTAGTAGAACAACCCCTGACCAAG	2025
DB	2108	GTATAGTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTGACATCCAAACGAAAG	2167
QY	2026	TCCACATCTTCAATGATGAGTGGGCAATGCCAGTGAAGATGTTGTTAGGCAAGAAAGCA	2085
DB	2168	TCCGCTATCAGATAATGTTTTCGTAACGCTAGTGACCACTGTTTCGTAATAAAGCAGA	2227

QY 2086 CAGTGAAGATCCAATAAGAACTTCAAGCGGATGAGAGCCAGTGTAGAGGAAACACCTGC 2145
 Db 2228 CCAAGATAGTAAACCTGATGAGATAGGAACATGATGAAGTAAAGTGAAGCCAACTCACCC 2287
 QY 2146 TGAGCCAGAAG 2156
 Db 2288 TGAATCTGATG 2298

RESULT 17
 US-10-158-844-243
 ; Sequence 243, Application US/10158844
 ; Publication No. US20040029118A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kunsch et al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-R
 ; COMPUTER: Dell Latitude Pentium 3
 ; OPERATING SYSTEM: Windows 98
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/158,844
 ; FILING DATE: 03-Jun-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/961,527
 ; FILING DATE: 1997-10-30
 ; APPLICATION NUMBER: US 60/029,960
 ; FILING DATE: 1996-10-31
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hyman, Mark J.
 ; REGISTRATION NUMBER: 46,789
 ; REFERENCE/DOCKET NUMBER: PB340P1D1
 ; INFORMATION FOR SEQ ID NO: 243:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2359 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 243:
 ; US-10-158-844-243

Query Match 27.4%; Score 653.8; DB 13; Length 2359;
 Best Local Similarity 68.3%; Pred. No. 6.6e-148;
 Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

QY 1 TTCCTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAAGGAAA---TAATCGTGTTC 57
 Db 939 TTCCTAGCTTGGAGCTTACCAAGCTGTGAGGATGAAGAGTCTTAATCGAGTTGC 998
 QY 58 CTATATAGATGGAACAAACAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGTAGAGTTAG 117
 Db 999 TTATATAGATGTGTATGATGAGCTGGTCAAAAGGCAGAAAACTTGACACCAAGATGAAGTCAG 1058
 QY 118 CAAGCGTGAAGAAATCAATGCTGAGCAATCGTATCAAGATAACAGACCAAGGCTATGT 177
 Db 1059 TAAAGAGGGAGGGGATCAACCGCGAACAAATTTGTTATCAAGATTAAGGATCAAGTTATGT 1118
 QY 178 CACTTCAATGCGGACCACTATCATATATTAATGATGATGATGATGATGATGATGATGATGAT 237
 Db 1119 GACCTCTCATGAGACCACTATCATATATTAATGATGATGATGATGATGATGATGATGATGAT 1178
 QY 238 CAGTGAAGAAATTAATCATGAAAGATCCAACTATAAGCTAAAAAGATGAGGATATTGTTAA 297

Db 1179 CAGTGAAGAGCTCTCATGAAGATCCGAATTTATCAGTTGAAGATTTCAGACATTTGTCAA 1238
 QY 298 TGAGGTCAAGGGTGCATATGTTATCAAGGTAGATGGAATACTACTATGTTTACCTTAAGGA 357
 Db 1239 TGAATCAAGGGTGTATGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGA 1298
 QY 358 TGCTGCCACGCGGATAACGTCCTACAAAGAGGAAATCAATCGACAAACAAAGAGCA 417
 Db 1299 TGCAGCTCATGCGGATAATATTCGGACAAAGAGAGATTAAACGTGAGAGAGGAGACG 1358
 QY 418 TAGTCAACATCGTGAAGGTGAACCTCAAGAAAGCATGGTCTGTTGCTTTGGCAGCTTC 477
 Db 1359 CAGTCATAATCAAACT-----CAAGAGCAGATAAATGCTGTTGTCGAGCGACAGC 1409
 QY 478 GCAAGGACGCTATACATACAGATGATGTTATATCTTTAATGCTTCTCATATCATAGAGGA 537
 Db 1410 CCAAGGAGGTTATACAACGGATGATGGTATATCTTCAATGATCTGATCATTTAGAGGA 1469
 QY 538 TACTGGTGTATATATATGCTTCTCATGAGATCATTTACCATTTACATTTCTTAAGAAATGA 597
 Db 1470 CACGGGTGATGCTTATATCTCTCACGCGACCATTTACCATTTACATTTCTTAAGAAATGA 1529
 QY 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTGCTGCGAGAAATCTGTC 657
 Db 1530 GTTATCAGCTAGCGAGTTAGCTGCTGCGAAGGCTATTGGAATGG----- 1574
 QY 658 AAATTCAAGAACCTATCGCGACAAAATAGCGATAACACTTTCAAGAACAACTGGGTACC 717
 Db 1575 -GAAGCAGGATCTCGTCTCTTCAAGTCTTAGTTATATGCAATCCAGCTCAACCAA 1633
 QY 718 TTCGTAAAGCAATCCAGGAATCAAAATACTAAACAAGCAACACAGCAAACTAACAG 777
 Db 1634 GATTGTCAAGAAACCAATCTGACTGT-----CACTCCAATTA 1673
 QY 778 TCAAGCAAGTCAAGTAATGACATTCATAGTCTCTTGAACAGCTCTCAAACTGCTTT 837
 Db 1674 TCATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTTAAACCTT 1733
 QY 838 GAGTCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGACAAATCAAGATCG 897
 Db 1734 ATCAGAACCCATGTGGATCTGATGGCTTTATTTGACCCAGCGCAATCAAGATCG 1793
 QY 898 AACAGCTAGAGGTGTGACAGTGCACACGAGATCATTTACCACTTCATCCCTTACTCTCA 957
 Db 1794 AACCGCCAGAGGTGTAGCTGTCCCTCATGGTAAACCATTTACCACTTTATCCCTTATGAACA 1853
 QY 958 AATGCTGAATTTGGAAGAGCAATCGCTGATATTTCCCTTCTGTTATCGTTTCAAAACA 1017
 Db 1854 AATGCTGAATTTGGAAGAAACGAATGCTGATATTTCCCTTCTGTTATCGTTTCAAAACA 1913
 QY 1018 TTGGGTACCAGATTCAGGCGCAGAAACAACTAGTCCACAAACGACTCCCGAACTAGTCC 1077
 Db 1914 TTGGGTACCAGATTCAGGCGCAGAAACCAAGTCCACAAACGACTCCAGAACTAGTCC 1973
 QY 1078 AGGCGCGCACTGACCAAAATCTTAAATAGACTCAAAATCTTTTGGTGTAGTCAAGT 1137
 Db 1974 AAGTCCGCAACACAGCTCCAAGCAATCCAATTTGATGAGAA-----ATTGGTCAAGAGAAGC 2027
 QY 1138 GGTACGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCGATCTCTCGTTATGTCCT 1197
 Db 2028 TGTTCGAAAGTAGGCGATGGTTATGTTTGGAGGAAATGGAGTTTCTCGTTATATCC 2087
 QY 1198 TGCGAAAGATTTACCATCTGAACTGTTAAATACTCTGAAAGCAAGTTATCAAAAACAGA 1257
 Db 2088 AGCCAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGATAGCAAACTGGCCACAGAGA 2147
 QY 1258 GAGTGTTCACACACTTTAACTGCTTAAAGAAATGTTGCTCTCGTGAACCAAGATTT 1317
 Db 2148 AAGTTTATCTCATAGCTAGGAACTTAAGAAACAGTCCCTCCCATCTAGTATCGAGAT 2207
 QY 1318 TTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAATTAAGGG 1377

Db 2208 TTACATAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACTTGCATATAAAGG 2267
 Qy 1378 TCCTAAATCTGATTTCCAGCCTTAGACAAATTTATTAGACGCTTGAATGATGATCGAC 1437
 Db 2268 TCACAGAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAAACGACTCAAGGATGTTCTCAAG 2327
 Qy 1438 TAATAAGAAATTTGATAGATTTTATGG 1469
 Db 2328 TGATAAAGTCAAGTTAGTGAAGATATTTCTTG 2359

RESULT 18
 US-09-765-272-181
 ; Sequence 181, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 181:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1342 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
 US-09-765-272-181

Query Match 16.1%; Score 385.4; DB 9; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 7.5e-83;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

Qy 37 TAAGGAAAATAATCGTGTTCCTATATAGATGGAACCAAGCAGCAGCAAAACCGAGAA 96
 Db 28 TAAGGACAAATAATCGTGTCTCTATGTGGATGGCAGCAGTCAAGTCAAGAAAGTGA 87
 Qy 97 TTTGACTCTGTAGAGTTAGCAAGCGTGAAGAAATCAATGCTGACCAATCGTCAATCA 156
 Db 88 CTTGACACAGACAGGTTAGCCAGAAAGAAATTCAGGCTGAGCAAAATGTAATCAA 147
 Qy 157 GATAACAGCAAGGTTATGCTACTTCCATGCGGACCACTATCAATTTATCAATGGTAA 216
 Db 148 AATTACAGATCAGGCTATGTAAGTCACACGCGTACCACTATCAATTTATTAATGGAA 207
 Qy 217 GGTTCCTTAGCGCTATCATCATGATGAGAAATTTACTCATGAAAGTCCAACTATAGCT 276
 Db 208 AGTTCTCTATGATGCCCTCTTTAGTGAAGAACTCTTTGATGAAGGATCCAAACTATCA 267

Qy 277 AAAAGATGAGGATATTTGTTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
 Db 268 TAAAGACGCTGATTTGTTCAATGAAGTCAAGGGTGGTTATATCATCAAGGTGATGAAA 327
 Qy 337 ATACTATGTTTACCTTAAAGATGCTGCCACGCGGATAAAGTCCGTACAAAGAGGAAT 396
 Db 328 ATATTATGCTACCTGAAAGATGAGCTCATGCTGATTAATGTTTGAAGTAAAGATGAAT 387
 Qy 397 CAATCGACAAAACAGAGCATATGTAACATCGTGAAGGTGGAACTCCAAGAAACGATGG 456
 Db 388 CAATCGTCAAAAACAGAACATGTCAAAGATAATGAGAAAGTTA-----ACTC 435
 Qy 457 TGCTGTTGCTTGGCAGCTTCGCAAGGACGCTTACTACAGATGATGGTTATATCTTTAA 516
 Db 436 TAATGTTGCTGTAGCAAGGCTTCAGGACGATATACGACAAATGATGTTATGTTCTTAA 495
 Qy 517 TGCTTCTGATATCATAGAGATCTGCTGATGCTTATATGTTCTCTCATGAGATCAATTA 576
 Db 496 TCCAGCTGATATATTCGAAGATACGGGTAATGCTTATGCTTCTCATGAGTCACTA 555
 Qy 577 CCATTACATTCCTAAGAAATGAGTTATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCT 636
 Db 556 TCCTACATTCCTCCAAAGCGATTTATCTGCTAGTGAATAGCAGCAGCTAAAGCACATCT 615
 Qy 637 ATCTGGTCAAGGAAATCTGCAAAATTCAGAACTTCAAGAACTTCCGCGACAAATAGCGATAC 696
 Db 616 GGCTGGAAATAATGCAACCGAGTCAAGTAAAGCTATTTCTTCAACAGCTAGTGACAA--- 672
 Qy 697 TCAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAATACTACTACACAAG 756
 Db 673 -----TAAACGCAATCTGT 687
 Qy 757 CAACAGACAGCAACTAAACAGTCAAGCAAGTCAAGTAATGACATTTGATGTTCTTTGAA 816
 Db 688 AGCAAAAGGATCAACTAGCAGCCAGCAATAAATCTGAAATCTCCAGAGTCTTTTGA 747
 Qy 817 ACAGCTCTACAACTGCTTTGAGTCAACGACATGTAGAAATCTGATGCTTGTCTTTGA 876
 Db 748 GGAATCTCTATGATTCACCTAGCGCCCAAGTTCAGATGAATCAGATGCGCTGCTTTGA 807
 Qy 877 TCCAGCAAAATCAAGTCCGAAGCAGTGAAGGTGTTTCAGTGCACACGAGATCAATTA 936
 Db 808 CCCTGCTAAGATTTACGTGCTACACCAATAGGAGTTCGATTCGCGATGGCAGCAATTA 867
 Qy 937 CCACCTTCTCTTACTCTCAAAATGCTGAAATTTGAAGAACGAATCGCTCGTATTTATCC 996
 Db 868 CCACCTTATCTTACAGCAAGCTTTCTGCTTGAAGAAAGATTGCCAGAAATGGTGCC 927
 Qy 997 CTTCTGT 1003
 Db 928 TATCAGT 934

RESULT 19
 US-09-769-787-246
 ; Sequence 246, Application US/09769787
 ; Publication No. US20030091577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21129WO
 ; CURRENT APPLICATION NUMBER: US/09/769,787
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: GB 9816337.1
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: US 60/125164
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 246

```

; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-246

Query Match      16.1%; Score 385.4; DB 10; Length 1455;
Best Local Similarity 65.3%; Pred. No. 7.9e-83;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCAAAAACCGGAA 96
Db 99 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCAAAAACCGGAA 158
QY 97 TTTGACTCTCTGATGAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 156
Db 159 CTTGACACCAAGCAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 218
QY 157 GATTAACAGCAAGCGTATGTCATCTCACATGGGACCACTATCATTTATTAATGGTAA 216
Db 219 AATTACAGATCAGGCTATGTAAGCTCACACGCTGACCACTATCATTTATTAATGGTAA 278
QY 217 GGTTCCTTATGACCTATCATGCTGAGGAATTAATCATGAAAGATCCAACTATAGCT 276
Db 279 AGTTCTTATGATCCCTCTTTAGTGAAGAACTCTTATGAGGATCCAACTATCACT 338
QY 277 AAAAGATGAGGATATGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
Db 339 TAAAGACGCTGATATTCATGAGTCAAGGTGAGGTGTTATATCATCAAGGTGATGAAA 398
QY 337 ATACTATGTTTACCTTAAGGATGCTGCCACGCGGATAAAGCTCGGTACAAAAGAGAAAT 396
Db 399 ATATTATGCTACCTGAAAGATGAGCTCATGCTGATAATGTTGCAACTTAAAGATGAAT 458
QY 397 CAATCGCAAAAACAAAGACATAGTCAACATCGTGAAGGTGGAAGTCCAAAGAACGATGG 456
Db 459 CAATCGTCAAAAACAAAGACATAGTCAACATCGTGAAGGTGGAAGTCCAAAGAACGATGG 506
QY 457 TGTCTGTGCTTGGCAGCGTTGCAAGGACGCTATACCTACAGATGATGGTTATATCTTTAA 516
Db 507 TAATGTTGCTGTAGCAAGGTCTCAGGACGATATACGCAAAATGATGGTTATGTTTAA 566
QY 517 TGCTTCTGATATCATGAGGATCTGCTGATGCTTATATCTGTTCTCTCATGGAGATCATTA 576
Db 567 TCCAGCTGATATATCGAAGATACGGGTAATGCTTATATGCTTCTCTCATGGAGTCACTA 626
QY 577 CCATTACATTCCTAAGAAATGAGTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 686
Db 627 TCACTACATTCCTAAGAAATGAGTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 686
QY 637 ATCTGCTGAGGAAATCTGTCAAAATCAAGAACTTACGCGGACAAATAGCGATACAC 696
Db 687 GGTGGAATAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA --- 743
QY 697 TTCAAGAACAACTGGGTACCTTCTGTGAAGAAATCAGGAATCTAACACAAG 756
Db 744 -----TAAACAGCAATCTGT 758
QY 757 CAACACAGCAACACTAACAGTCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 816
Db 759 AGCAAAAGATCACTTAGCAAGCAGCAATTAATCTGAAATCTCCAGAGCTTTTGA 818
QY 817 ACAGCTCTCAAACTGCTTTGAGTCAAGCAATGATGAAATCTGATGGCTTGTCTTTGA 876
Db 819 GGAACCTTATGATTCACCTAGGCCCAAGCTTACAGTGAATCAGATGGCTTGTCTTTGA 878
QY 877 TCCAGCAAAATCAAGTCAAGCTAGAGGTGTTGAGTGGCCACACGAGATCATTA 936
Db 879 CCCTGCTAAGATTTATCAGTCTGACCAAAATGAGTTGCGATTCGCGATGGCGACCATTA 938
QY 937 CCACCTTATCCCTTACTCTCAATGCTGAAATGGNAGAACGAATCGCTGATTTATCCC 996
Db 939 CCACCTTATTCCTTACGCAAGCTTTCTGCTGCTTAGAAGAAAGATTTGCCAGAAATGGTGC 998

```

```

QY 997 CTTTCGT 1003
Db 999 TATCAGT 1005

RESULT 20
US-09-769-744A-23
; Sequence 23, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIORITY FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-23

```

```

Query Match      16.1%; Score 385.4; DB 10; Length 1455;
Best Local Similarity 65.3%; Pred. No. 7.9e-83;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCAAAAACCGGAA 96
Db 99 TAAGGAAATATCGTGTTCCTATATAGATGGAACAAACGCGACGCAAAAACCGGAA 158
QY 97 TTTGACTCTCTGATGAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 156
Db 159 CTTGACACCAAGCAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCA 218
QY 157 GATTAACAGCAAGCGTATGTCATCTCACATGGGACCACTATCATTTATTAATGGTAA 216
Db 219 AATTACAGATCAGGCTATGTAAGCTCACACGCTGACCACTATCATTTATTAATGGTAA 278
QY 217 GGTTCCTTATGACCTATCATGCTGAGGAATTAATCATGAAAGATCCAACTATAGCT 276
Db 279 AGTTCTTATGATCCCTCTTTAGTGAAGAACTCTTATGATGAAGGATCCAACTATCACT 338
QY 277 AAAAGATGAGGATATGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA 336
Db 339 TAAAGACGCTGATATTCATGAGTCAAGGTGAGGTGTTATATCATCAAGGTGATGAAA 398
QY 337 ATACTATGTTTACCTTAAGGATGCTGCCACGCGGATACGCTCCGTACAAAAGAGAAAT 396
Db 399 ATATTATGCTACCTGAAAGATGAGCTCATGCTGATATGTTGCAACTTAAAGATGAAT 458
QY 397 CAATCGCAAAAACAAAGACATAGTCAACATCGTGAAGGTGGAAGTCCAAAGAACGATGG 456
Db 459 CAATCGTCAAAAACAAAGACATAGTCAACATCGTGAAGGTGGAAGTCCAAAGAACGATGG 506
QY 457 TGTCTGTGCTTGGCAGCGTTGCAAGGACGCTTACTACAGATGATGGTTATATCTTTAA 516
Db 507 TAATGTTGCTGTAGCAAGGTCTCAGGACGATATACGCAAAATGATGGTTATGTTTAA 566
QY 517 TGCTTCTGATATCATGAGGATCTGCTGATGCTTATATCTGTTCTCTCATGGAGATCATTA 576
Db 567 TCCAGCTGATATATCGAAGATACGGGTAATGCTTATATGCTTCTCTCATGGAGTCACTA 626
QY 577 CCATTACATTCCTAAGAAATGAGTTATCTGCTAGTGAATTAGCAGCAGCTTAAAGCATCT 636

```

Db	627	TCACATCTCCCAAGAGGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCT	686
Qy	637	ATCTGGTCGAGGAATCTGCAAAATCAAGAACTTATCGCCGGACAAAATAGCGATAACAC	696
Db	687	GGCTGGAAAAAATATGCAACCGAGTCAGTTAAAGCTATTTCTTCAACAGCTAGTGACAA---	743
Qy	697	TTCAAGAACAACTGGGTACCTTCTGTAAGCAATCCAGGAACTACAATACTAACACAAG	756
Db	744	-----TAAACGCAATCTGT	758
Qy	757	CAACACAGCAACACTAACAGCTCAAGCAAGTCAAAAGTAATGACATGATCTCTTGAA	816
Db	759	AGCAAGAGGATCACTAGCAGCCAGCAAAATAATCTCGAATGGCGACCATTA	818
Qy	817	ACAGCTTACAAAATGCTTGAATGCTGAAATGGGAAGCAATCGCTCGTATATATTC	876
Db	819	GGAACTCTATGATTCACCTAGGCGCCACGTTACAGTGAATCAGATGGCTGGTCTTTGA	878
Qy	877	TCCAGCAAAATCAAGTCAAGCTAGAGGTGAGTGAAGCAATGCTGAGCAATCTGATCA	936
Db	879	CCCTGCTAAGATTTATCAGTCGTACACCAATGGAGTTGCGATTCGCGATGGCGACATTA	938
Qy	937	CCACTTCATCCCTTACTCTCAATGCTGAAATGGGAAGCAATCGCTCGTATATATTC	996
Db	939	CCACTTTATCTCTTACAGCAAGCTTTCTGCTTAGAAGAAAGATTCAGCAATGGTGCC	998
Qy	997	CCTTCGT 1003	
Db	999	TATCAGT 1005	
RESULT 21			
US-10-412-862-7			
; Sequence 7, Application US/10412862			
; Publication No. US20040052781A1			
; GENERAL INFORMATION:			
; APPLICANT: Johnson, Leslie S.			
; APPLICANT: Adamou, John E.			
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural			
; TITLE OF INVENTION: Motifs			
; FILE REFERENCE: 469201-685			
; CURRENT APPLICATION NUMBER: US/10/412,862			
; PRIOR FILING DATE: 2003-04-14			
; PRIOR APPLICATION NUMBER: 09/468,656			
; PRIOR FILING DATE: 1999-12-21			
; PRIOR APPLICATION NUMBER: 60/113,048			
; PRIOR FILING DATE: 1998-12-21			
; NUMBER OF SEQ ID NOS: 14			
; SOFTWARE: Patent In Ver. 3.0			
; SEQ ID NO 7			
; LENGTH: 1455			
; TYPE: DNA			
; ORGANISM: Streptococcus pneumoniae			
US-10-412-862-7			
Query Match			
Best Local Similarity 16.1%; Score 385.4; DB 13; Length 1455;			
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;			
Qy	37	TAAAGAAAATAATCGTGTTCCTATATAGATGAAACAAAGCAGCAGCAAAAACGGAGAA	96
Db	99	TAAAGCAATATCGTGTCTCTTATGATGCGCAGCTCAAGTCAGAAAAGTGAATA	158
Qy	97	TTTGACTCTGATGAGGTAGCAAGCGTGAAGAAATGCTGAGCAATCTGATCA	156
Db	159	CTTGACACCAAGCAGGTAGCCAGAAAGAAATTCAGGCTGAGCAATTTGATCA	218
Qy	157	GATAACAGACCAAGGCTATCTCACTTACATGCGACCATCTATCTATTAATGATA	216
Db	219	AATTACAGATCAGGCGCTATGATTAACGTCACACGCTGACCATCTATCTATGATA	278
Qy	217	GGTTCTTATGACGCTATCATCTAGTGAAGATTTACTCATGAAGATCCAAACTATAAGCT	276
Db	279	AGTTCTTATGATGCCCCCTTTAGTGAAGAACTCTTATGATGAAGATCCAAACTATCACT	338
Qy	277	AAAAGATGAGGATATCTGTAATGAGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	336
Db	339	TAAAGCGCTGATATGTCATGAGTCAAGGGTGGTATATCATCAAGGTGATGAAA	398
Qy	337	ATACTATGTTTACCTTAAGGATGCTCCCAAGCGGATPAACGTCCGTCAAAAGAGGAAT	396
Db	399	ATATTATGTCACCTGAAAGATGAGCTCATCTGCTGATAATGTTCCAACTAAAGATGAAT	458
Qy	397	CAATCGACAAAACAGAGCATAGTCAACATCGTGAAGGTGGAATCCAGAAACGATGG	456
Db	459	CAATCGTCAAAAACAGAACATGTCAAGATAATGAGAAGTTA-----ACTC	506
Qy	457	TGCTGTGCTTGCACCTTCGCAAGGACGCTATCTATACAGATGATGTTATATCTTTAA	516
Db	507	TAATGTTGCTGACCAAGTCTCAGGACGATATACGACAAATGATGTTATGTTTAA	566
Qy	517	TGCTTCTGATATCATAGAGGATCTGGTGAATGCTTATATCTGCTCATGAGATCATTA	576
Db	567	TCCAGCTGATATTCGAAGATACGGGTAAATGCTTATATCTGCTCATGAGGTCACTA	626
Qy	577	CCATTACATTCCTAAGATGATTTATCAGCTAGCGAGTTGGCTGCTGAGAAAGCCTTCT	636
Db	627	TCCTACATTTCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAGACATCT	686
Qy	637	ATCTGCTGAGGAAATCTGTCAAAATTCAGAAACCTATCGCGACAAAATAGCGATAACAC	696
Db	687	GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAA---	743
Qy	697	TTCAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAACTACAATACTAACACAAG	756
Db	744	-----TAAACGCAATCTGT	758
Qy	757	CAACACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAATGACATGATCTCTTGAA	816
Db	759	AGCAAGAGGATCACTAGCAGCCAGCAAAATAATCTCGAATGGCGACCATTA	818
Qy	817	ACAGCTTACAAAATGCTTGAATGCTGAAATGGGAAGCAATCGCTCGTATATATTC	876
Db	819	GGAACTCTATGATTCACCTAGGCGCCACGTTACAGTGAATCAGATGGCTGGTCTTTGA	878
Qy	877	TCCAGCAAAATCAAGTCAAGCTAGAGGTGAGTGAAGCAATGCTGAGCAATCTGATCA	936
Db	879	CCCTGCTAAGATTTATCAGTCGTACACCAATGGAGTTGCGATTCGCGATGGCGACATTA	938
Qy	937	CCACTTCATCCCTTACTCTCAATGCTGAAATGGGAAGCAATCGCTCGTATATATTC	996
Db	939	CCACTTTATCTCTTACAGCAAGCTTTCTGCTTAGAAGAAAGATTCAGCAATGGTGCC	998
Qy	997	CCTTCGT 1003	
Db	999	TATCAGT 1005	
RESULT 22			
US-10-412-850-7			
; Sequence 7, Application US/10412850			
; Publication No. US20040001836A1			
; GENERAL INFORMATION:			
; APPLICANT: Johnson, Leslie S.			
; APPLICANT: Adamou, John E.			
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus			
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural			
; TITLE OF INVENTION: Motifs			
; FILE REFERENCE: 469201-686			
; CURRENT APPLICATION NUMBER: US/10/412,850			
; PRIOR FILING DATE: 2003-04-14			
; PRIOR APPLICATION NUMBER: 09/468,656			
; PRIOR FILING DATE: 1999-12-21			
; PRIOR APPLICATION NUMBER: 60/113,048			

; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-850-7

Query Match 16.1%; Score 385.4; DB 16; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 7.9e-83;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY	37	TAAGGAAATAATCGTGTTCCTATATAGATGGAAAAAAGAGCGACGCAAAAAACGGAGAA	96
Db	99	TAAGGACAATAATCGTGTCTCTTATGTGATGGCAGCCAGTCAAGTCAGAAAGTGAAAA	158
QY	97	TTTGACTCTGATGAGTTAGCAAGGGTGAAGAAATCAATGTGAGCAAAATCGTCATCAA	156
Db	159	CTTGACACGACACAGGTTAGCCAGAAAAGAGAAATTCAGGCTGAGCAAAATGTAATCAA	218
QY	157	GATAACAGACCAAGCTATGTCACCTTCACATGGCGGACCACTATCAATTATCAATGGTAA	216
Db	219	AATTACAGATCAGGCTAATGTAAGTCAACGTCACGTCAGCCACTATCAATTATTAATGGAA	278
QY	217	GGTTCCTTATGACGCTATCATCAGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	276
Db	279	AGTTCTCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGATCCAAATATCAACT	338
QY	277	AAAAGATCAGATATTTGTAAGTCAAGGGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	336
Db	339	TAAAGACGCTGATATTTGTAAGTCAAGGGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	398
QY	337	ATACTATGTTTACCTTAAGGATGCTCCCAACGGGATACGTTCCGTACAAAAGAGAAAT	396
Db	399	ATATTATGTTACCTGAAAGATGACGCTCATGCTGATATGTTTGAAGTAAAGATGAAAT	458
QY	397	CAATCGTCAAAAACCAAGAGCATGTCAAGATTAATGAGAAGTTA-----ACTC	506
Db	507	TGCTGTGCTTGGCAGCTTCGCAAGGACGCTATACAGATGATGGTTATATCTTTAA	516
QY	507	TAATGTTGCTGTAGCAAGGCTCTCAGGAGCATATACGACAAATGATGTTATGCTTTAA	566
QY	517	TGCTCTGATATCATAGAGGATACGTTGATGCTTATATCGTTTCTCATGGAGATCAATTA	576
Db	567	TCAGCTGATATTCGGAAGATACGGGTAATGCTTATATCGTTTCTCATGGAGGTCACCTA	626
QY	577	CAATTACATTCCTAAGAAATGAGTTATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCCT	636
Db	627	TCACATATTCCTCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAGCACAATCT	686
QY	637	ATCTGTCGAGGAAATCTGTCAATTCAGAACCTATCGCCGCAAAATAGCGATACAC	696
Db	687	GGCTGGAATAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA--	743
QY	697	TTCAAGAACAAACTGGGTACCTTCTGTAAGCAATCCAGGAATCAAAATATCAACAAG	756
Db	744	-----TAACACGCAATCTGT	758
QY	757	CAACAACAGCAACCTAACAGTCAAGCAAGTCAAAAGTGAATGATGATGATCTCTTGAA	816
Db	759	AGCAAAAGGATCACTAGCAAGCGACGCAAAATAATCTGAAATCTCCAGAGCTCTTTGA	818
QY	817	ACAGCTCTCAAACTGCTTTGAGTCAACGACATGTAGAAATCTGATGGCTTGTCTTTGA	876
Db	819	GGAACTCTTATGATTCACCTACGTCGACCAAGCTTACAGTGAATCAGATGGCTGCTTTGA	878
QY	877	TCAGACACAAATCACAAGTCGACAGCTAGAGGTTGAGTGTGAGTGCACACAGGAGATCAATTA	936
Db	879	CCCTGCTAAGATTTACAGTCGTACCAAAATGAGGTTGGGATTCGCGATGGGACCAATTA	938

QY	937	CCACTTCATCCCTTACTCTCAAAATGCTGAATTTGAAGAACAATCGTCGTATTATTC	996
Db	939	CCACTTTTATTCCTTACTACAGCAAGCTTTCTGCTTTAGAAGAAAAAGATTGCCAGATGCTGCC	998
QY	997	CCTTCGT 1003	
Db	999	TATCAGT 1005	

RESULT 23

US-10-387-783-7
 ; Sequence 7, Application US/10387783
 ; Publication No. US2004000531A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John B.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-683
 ; CURRENT APPLICATION NUMBER: US/10/387,783
 ; CURRENT FILING DATE: 2003-03-13
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-387-783-7

Query Match 16.1%; Score 385.4; DB 16; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 7.9e-83;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY	37	TAAGGAAATAATCGTGTTCCTATATAGATGGAAAAAAGAGCGACGCAAAAAACGGAGAA	96
Db	99	TAAGGACAATAATCGTGTCTCTTATGTGATGGCAGCCAGTCAAGTCAGAAAGTGAAAA	158
QY	97	TTTGACTCTGATGAGTTAGCAAGGGTGAAGAAATCAATGTGAGCAAAATCGTCATCAA	156
Db	159	CTTGACACGACACAGGTTAGCCAGAAAAGAGAAATTCAGGCTGAGCAAAATGTAATCAA	218
QY	157	GATAACAGACCAAGCTATGTCACCTTCACATGGCGGACCACTATCAATTATTAATGGTAA	216
Db	219	AATTACAGATCAGGCTAATGTAAGTCAACGTCACGTCAGCCACTATCAATTATTAATGGAA	278
QY	217	GGTTCCTTATGACGCTATCATCAGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	276
Db	279	AGTTCTCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGATCCAAATATCAACT	338
QY	277	AAAAGATGAGATATTTGTAAGTCAAGGGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	336
Db	339	TAAAGACGCTGATATTTGTAAGTCAAGGGTGAAGAAATTAATCTGTAAGAAATCCAAATATAAGCT	398
QY	337	ATACTATGTTTACCTTAAGGATGCTCCCAACGGGATACGTTCCGTACAAAAGAGAAAT	396
Db	399	ATATTATGTTACCTGAAAGATGACGCTCATGCTGATATGTTTGAAGTAAAGATGAAAT	458
QY	397	CAATCGTCAAAAACCAAGAGCATGTCAAGATTAATGAGAAGTTA-----ACTC	506
QY	457	TGCTGTGCTTGGCAGCTTCGCAAGGACGCTATACAGATGATGGTTATATCTTTAA	516
Db	507	TAATGTTGCTGTAGCAAGGCTCTCAGGAGCATATACGACAAATGATGTTATGCTTTAA	566
QY	517	TGCTCTGATATCATAGAGGATACGTTGATGCTTATATCGTTTCTCATGGAGATCAATTA	576
Db	567	TCAGCTGATATTCGGAAGATACGGGTAATGCTTATATCGTTTCTCATGGAGGTCACCTA	626

QY 577 CCATTACATCTCTAAGAAATGAGTTATCAGCTAGCGAGTTGGCTCTCGAGAGCCTTCCT 636
 Db 627 TCACTACATCTCCCAAGCGAATTTATCTGCTAGTGAATTTAGCAGCAGCTAAGCACAATCT 686
 QY 637 ATCTGGTCGAGGAATCTGTCAAAATTCAGAACTATCGCGACAAAATACGGATAACAC 696
 Db 687 GGCTGGAAAAAATATGCAACCGAGTCAAGTTAAGCTATTCCTCAACAGCTAGTGAACA 743
 QY 697 TTCAAGAACAACTGGGTACCTCTGTAGCAATCCAGAACTACAAATACATAACACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 QY 757 CAACAACAGCAACACTAAGCTAGCAAGTCAAGAAATGAGTATGATCTCTTTGAA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAATCTCCAGAGCTCTTTGAA 818
 QY 817 ACAGCTCTACAACTGGCTTTGAGTCAAGCAATGTAAGTCTGATGCTCTGCTTTGA 876
 Db 819 GGAATCTATGATTCACCTAGGCGCAAGCTTACAGTGAATCAGATGCTGCTTTGA 878
 QY 877 TCCAGCAAAATCAAGTCAAGCTAGAGGTTGCGAGTGCACACGAGATCAATTA 936
 Db 879 CCCTGCTAAGATTTACGTGTACACCAATGAGGTTGCGATTCGCGATGGCGACCAITTA 938
 QY 937 CCATTCATCCCTTACTCTCAATGTCTGAATTTGGAAGCAATCGCTCGTATTATTC 996
 Db 939 CCACCTTTATCTCTACAGCAAGCTTTCTGCTTTAGAAAGAAAGATTGCCAGAAATGGTGCC 998
 QY 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 24
 US-09-884-465A-1
 ; Sequence 1, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 394
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 3120
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-884-465A-1
 Query Match 16.1%; Score 385.4; DB 10; Length 3120;
 Best Local Similarity 65.3%; Pred. No. 1.2e-82;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAGGAAATATATCGTGTCTTTCTATATAGTGAAGAAATCAAGCGAGCGCAAAACGGAGAA 96
 Db 99 TAAGGAAATATATCGTGTCTTTCTATGTTGGATGGCAGCAGTCAAGTCAGAAAGTGA 158
 QY 97 TTTGACTCTGATGAGCTAGCAGCTGAGGAAATCAATGCTGAGCAATCGTCATCAA 156
 Db 159 CTTGACACCAAGCAGGTTAGCCAGAAAGAGGAAATTCAGGCTGAGCAATTTGAATCAA 218
 QY 157 GATAACAGACCAAGGCTATGTCATTCACATGGCGACCATATCATTTATCAATGGTAA 216

Db 219 AATTACAGATCAGGGCTATGTAAAGCTCACGGTGACCACTATCATTTACTATATGGAA 278
 QY 217 GGTTCCTTTATGACCTATCATCTAGTGAAGAAATTTACTCATGAAAGATCCAAACTATAGCT 276
 Db 279 AGTTCTTTATGATGCCCTCTTTAGTGAAGAACTCTTGTATGAAGGATCCAAACTATCACT 338
 QY 277 AAAAGATGAGGATATTTAATGAGGTCAGGGTGGATATGTTATCAAGGTAGTAGGAA 336
 Db 339 TAAAGACGCTGATATTCTCAATGAAGTCAAGGGTGGTTATATCATCATCAAGGTCGATG 398
 QY 337 ATACTATGTTTACCTTAAGGATGCTCCCGCGGATAACGTCCTGTAACAAAGAGGAAAT 396
 Db 399 ATATTATGCTTACCTGAAGATGCGCTCATGCTGATATGTTTCAACTAAGATGAAT 458
 QY 397 CAATCGCAAAAAAACAAGAGCATAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGG 456
 Db 459 CAATCGTCAAAAAACAAGACATGTCNAAGATAATAGAGAAGTTA-----ACTC 506
 QY 457 TGCTGTTGCCCTGGCAGCTTCGAAAGGCGCTTACTACAGATGATGTTTATATCTTTAA 516
 Db 507 TAATGTTGCTCTAGCAAGGCTCTCAGGAGCATATACGACAAATGATGTTTATGCTTTAA 566
 QY 517 TGCTTCTCATATCATAGAGGATGCTGCTGATGCTTTATATCGTTCTCATGAGATCAATTA 576
 Db 567 TCCAGCTGATATTTATCGAAGATACCGGTAATGCTTTATCGTTCTCATGAGGTCATTA 626
 QY 577 CCATTACATCTCTAAGAAATGATTTATCAGCTAGCGAGTGGCTGTCGAGAAAGCCTTCCT 636
 Db 627 TCACCTACATTTCCAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAGACATCT 686
 QY 637 ATCTGTCGAGGAATCTGTCAATTCAGAACCTATCGCCGACAAAATAGGATAACAC 696
 Db 687 GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAA 743
 QY 697 TTCAAGAACAAACTGGTACCTTCTGTAGCAATCAGGAACCTACAAATACTAACAACAAG 756
 Db 744 -----TAAACGCAATCTGT 758
 QY 757 CAACAACAGCAACACTAAGCTCAAGCAAGTCAAGTAAAGTAAAGTATGATGCTCTTTGAA 816
 Db 759 AGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAATCTCCAGAGTCTTTGAA 818
 QY 817 ACAGCTCTCAAACTGCTTTGAGTCAAGCATGATGATCTGATGCTGCTGCTTTGCA 876
 Db 819 GGAATCTATGATTCACCTAGGCGCAAGCTTACAGTGAATCAGATGCTGCTGCTTTGA 878
 QY 877 TCCAGCAAAATCAAGTCAAGCTAGAGGTTGCGAGTGCACACGCGAGATCAATTA 936
 Db 879 CCCTGCTAAGATTTACGTCGTACACCAATGAGTGGATTCGCGATGGCGACCAITTA 938
 QY 937 CCACCTTATCTCTTACAGCAAGCTTTCTGCTTTAGAAAGAAAGATTGCCAGAAATGGTGCC 998
 Db 997 CCTTCGT 1003
 Db 999 TATCAGT 1005

RESULT 25
 US-09-884-465A-2
 ; Sequence 2, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A

; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 5048
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-2

Query Match 16.1%; Score 385.4; DB 10; Length 5048;
 Best Local Similarity 65.3%; Pred. No. 1.6e-82;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY	37	TAAGGAAATATCGTGTTCCTATATAGTGAAGAAACAGCGACGCAAAAACGGAGAA	96
Db	1875	TAAGGACATATATCGTGTTCCTATATAGTGAAGAAACAGCGACGCAAAAACGGAGAA	1934
QY	97	TTTGACTCCTGATGAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAA	156
Db	1935	CTTGACACACAGCAGGTAGCAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAA	1994
QY	157	GATACAGACCAAGCCTATGTCATTCACATGGCGACCACTATCATTTATTAATGGTAA	216
Db	1995	AATTACAGATCAGGCGCTATGTAACGTGACACGGTGACCACTATCATTTATTAATGGTAA	2054
QY	217	GGTTCCTTATGACGCTATCATGCTGAGGAATTAATGCTGAGCAAAATCGTCATCAA	276
Db	2055	AGTTCCTTATGATGCGCTTATGTAAGGAATTAATGCTGAGCAAAATCGTCATCAA	2114
QY	277	AAAAGATGAGATATGTTAATAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	336
Db	2115	TAAAGACGCTGATATGTTAATAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	2174
QY	337	ATACTATGTTTACCTTAAGGATGCTGCCACGCGGATAAGCTCCGTACAAAAGAGAAAT	396
Db	2175	ATATTATGTTTACCTTAAGGATGCTGCCACGCGGATAAGCTCCGTACAAAAGAGAAAT	2234
QY	397	CAATCGCAAAAAACAGAGATAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGG	456
Db	2235	CAATCGTCAAAAAACAGAGATAGTCAACATCGTGAAGGTGGAACTCCAGAAACGATGG	2282
QY	457	TGCTGTTCCTTGGCAGTTCGCAAGGACGCTATATACAGATGATGGTTATATCTTTAA	516
Db	2283	TAATGTTCCTTGGCAGTTCGCAAGGACGCTATATACAGATGATGGTTATATCTTTAA	2342
QY	517	TGCTTCCTGATATCATAGAGATAGTGGTATGCTTATATCGTTCCTCATCGAGATCATTA	576
Db	2343	TCAGCTGATATATCATAGAGATAGTGGTATGCTTATATCGTTCCTCATCGAGATCATTA	2402
QY	577	CCATTACATTCCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	636
Db	2403	TCATTACATTCCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2462
QY	637	ATCTGTTCGAGGAATCTGTCAAATTCAGAACTATCGCCGACAAAATAGCGATAACAC	696
Db	2463	GGCTGGAATAAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA	2519
QY	697	TTCAGAACAACTGGGTACCTTCTGTAAGCAATCAGGAACTACAAATACTAACACAAG	756
Db	2520	-----TAACACGCAATCTGT 2534	
QY	757	CAACACAGCAACACTAACAGTCAAGCAAGTCAAGCAAGTCAAGCAAGTCAAGCAAGT	816
Db	2535	AGCAAAAGGATCAACTAGCAAGCAGCAAAATATCTGAAAATCTCCAGAGCTTTTGA	2594
QY	817	ACAGCTCTCAAACTCCCTTTGAGTCAACGACATGATGATGATGATGATGATGATGATGAT	876
Db	2595	GGAACCTCTATGATTCACCTAGCGCCCAACGTTACAGTGAATCAGATGGCTGTCTTGA	2654
QY	877	TCCAGCACAAATCAGAGTCGAACTAGAGGTGTTGAGTGGCCACACGAGATCATTA	936

Db	2655	CCCTGCTAAGATTATCAGTGTGATACCAAAATGGAGTTGGATTCCGCAATGGGACCATTA	2714
QY	937	CCACTTCATCCCTTACTCTCAAAATGCTGAATTTGAAGAAACGAATCGCTGATTTATTC	996
Db	2715	CCACTTTATTCCTTACAGCAAGCTTTCTGCTTTAGAGAAAGATTGCCAGATGGTGCC	2774
QY	997	CCTTCGT 1003	
Db	2775	TATCAGT 2781	

Search completed: September 30, 2004, 22:39:20
 Job time : 1157 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: October 1, 2004, 03:33:09 ; Search time 81 Seconds
 (without alignments)
 2776.640 Million cell updates/sec

Title: US-09-765-271-56
 Perfect score: 4165
 Sequence: 1 SYELGLQARTVKENRRVSY.....KILALLKGNSPSSVSKKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1596107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1596107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

- 1: Geneseqp29Jan04.*
- 2: Geneseqp1980s.*
- 3: Geneseqp1990s.*
- 4: Geneseqp2000s.*
- 5: Geneseqp2001s.*
- 6: Geneseqp2002s.*
- 7: Geneseqp2003as.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	ID	Description
1	4163	100.0	796	2	AAW55090 Streptoco
2	4163	100.0	796	5	Abp54584 S. pneumo
3	4163	100.0	796	6	Adc45137 S. pneumo
4	4163	100.0	802	6	Abu01598 S. pneumo
5	4163	100.0	819	3	AAU01468 Recombina
6	3218	77.3	821	3	AAU01468 Recombina
7	3218	77.3	821	3	AAU01468 Recombina
8	3218	77.3	821	5	AAU01468 Recombina
9	3218	77.3	821	7	AAU01468 Recombina
10	3218	77.3	840	5	AAU01468 Recombina
11	3218	77.3	840	5	AAU01468 Recombina
12	3218	77.3	840	5	AAU01468 Recombina
13	2961	71.1	690	3	AAU01468 Recombina
14	2961	71.1	690	3	AAU01468 Recombina
15	2961	71.1	690	3	AAU01468 Recombina
16	2812	67.5	826	3	AAU01468 Recombina
17	2797.5	67.2	819	6	AAU01468 Recombina
18	2795	67.1	827	3	AAU01468 Recombina
19	2788.5	67.0	819	3	AAU01468 Recombina
20	2786	66.9	805	3	AAU01468 Recombina
21	2772	66.6	820	3	AAU01468 Recombina
22	2772	66.6	838	3	AAU01468 Recombina
23	2772	66.6	839	6	AAU01468 Recombina
24	2771	66.5	807	3	AAU01468 Recombina
25	2760	66.3	811	3	AAU01468 Recombina

26	2750.5	66.0	819	3	AAU01468 Recombina
27	2750.5	66.0	819	3	AAU01468 Recombina
28	2750.5	66.0	819	5	AAU01468 Recombina
29	2750.5	66.0	819	7	AAU01468 Recombina
30	2750.5	66.0	838	3	AAU01468 Recombina
31	2750.5	66.0	838	5	AAU01468 Recombina
32	2750.5	66.0	838	7	AAU01468 Recombina
33	2733	65.6	834	3	AAU01468 Recombina
34	2728	65.5	816	3	AAU01468 Recombina
35	2727	65.5	816	3	AAU01468 Recombina
36	2720	65.3	811	3	AAU01468 Recombina
37	2720	65.3	811	3	AAU01468 Recombina
38	2714	65.2	816	3	AAU01468 Recombina
39	2713	65.1	811	3	AAU01468 Recombina
40	2649.5	63.6	763	2	AAU01468 Recombina
41	2649.5	63.6	763	5	AAU01468 Recombina
42	2649.5	63.6	763	7	AAU01468 Recombina
43	2431.5	58.4	721	2	AAU01468 Recombina
44	2170	52.1	613	3	AAU01468 Recombina
45	2170	52.1	613	5	AAU01468 Recombina
46	2170	52.1	613	7	AAU01468 Recombina
47	1898	45.6	1126	5	AAU01468 Recombina
48	1898	45.6	1126	7	AAU01468 Recombina
49	1891.5	45.4	1365	5	AAU01468 Recombina
50	1891.5	45.4	1365	7	AAU01468 Recombina
51	1888	45.3	555	3	AAU01468 Recombina
52	1888	45.3	555	5	AAU01468 Recombina
53	1888	45.3	555	7	AAU01468 Recombina
54	1808	43.4	612	5	AAU01468 Recombina
55	1808	43.4	612	7	AAU01468 Recombina
56	1795.5	43.1	568	3	AAU01468 Recombina
57	1795.5	43.1	568	5	AAU01468 Recombina
58	1795.5	43.1	568	7	AAU01468 Recombina
59	1795.5	43.1	1139	5	AAU01468 Recombina
60	1795.5	43.1	1139	7	AAU01468 Recombina
61	1795.5	43.1	1378	5	AAU01468 Recombina
62	1795.5	43.1	1378	7	AAU01468 Recombina
63	1666	40.0	334	3	AAU01468 Recombina
64	1666	40.0	334	5	AAU01468 Recombina
65	1666	40.0	334	7	AAU01468 Recombina
66	1661	39.9	1238	5	AAU01468 Recombina
67	1661	39.9	1238	7	AAU01468 Recombina
68	1655	39.7	999	5	AAU01468 Recombina
69	1655	39.7	999	7	AAU01468 Recombina
70	1651	39.6	999	5	AAU01468 Recombina
71	1651	39.6	999	7	AAU01468 Recombina
72	1649	39.6	428	3	AAU01468 Recombina
73	1649	39.6	428	5	AAU01468 Recombina
74	1649	39.6	428	7	AAU01468 Recombina
75	1614.5	38.8	473	3	AAU01468 Recombina
76	1614.5	38.8	473	5	AAU01468 Recombina
77	1614.5	38.8	473	7	AAU01468 Recombina
78	1591.5	38.2	1057	3	AAU01468 Recombina
79	1591.5	38.2	1058	5	AAU01468 Recombina
80	1591.5	38.2	1058	7	AAU01468 Recombina
81	1552	37.3	487	3	AAU01468 Recombina
82	1552	37.3	487	5	AAU01468 Recombina
83	1552	37.3	487	7	AAU01468 Recombina
84	1552	37.3	487	5	AAU01468 Recombina
85	1552	37.3	487	7	AAU01468 Recombina
86	1247	29.9	679	5	AAU01468 Recombina
87	1247	29.9	780	3	AAU01468 Recombina
88	1247	29.9	780	7	AAU01468 Recombina
89	1247	29.9	1019	3	AAU01468 Recombina
90	1247	29.9	1019	3	AAU01468 Recombina

ALIGNMENTS

RESULT 1

AAW55090

ID AAW55090 standard; protein; 796 AA.

XX AAW55090;
AC 02-OCT-1998 (first entry)
DT Streptococcus pneumoniae SP0036 protein.
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
DE Streptococcus pneumoniae; otitis media; meningitis.
XX Streptococcus pneumoniae.
OS
XX Key Location/Qualifiers
FH Misc-difference 456
FT /label= unknown
FT /note= "encoded by GNA"
XX WO9818930-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US019422.
XX 31-OCT-1996; 96US-0029960P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX WPI; 1998-272224/24.
DR N-PSDB; AAV27351.
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 59-60; 118pp; English.
PS
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
XX Sequence 796 AA;
SQ
Query Match 100.0%; Score 4163; DB 2; Length 796;
Best Local Similarity 100.0%; Pred. No. 9.4e-299;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRNSVSDGQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLYQARTVKNRNSVSDGQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNPKYKDEDIVNEVKGKGYVUKVDGKYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISEELMKDPNPKYKDEDIVNEVKGKGYVUKVDGKYVYLKDA 120
QY 121 AHADNVTKEINPKQKHSQHRREGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
Db 121 AHADNVTKEINPKQKHSQHRREGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDT 180
QY 181 GDYIVPHGDHYHYIPKNELASASELAFAFLSGRGNLSNSRTYRRQNSDNTSRTNWPS 240

Db 181 GDYIVPHGDHYHYIPKNELASASELAFAFLSGRGNLSNSRTYRRQNSDNTSRTNWPS 240
QY 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQLYKLPLSQRHVESDGLVDFDPAQITSRT 300
Db 241 VSNPGTNTNTNSNTNSQASQSDIDSLKQLYKLPLSQRHVESDGLVDFDPAQITSRT 300
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERLARIIPLYRSNHVWVDSRPPSPQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHYIPYSQMSLEERLARIIPLYRSNHVWVDSRPPSPQPTPEPSPG 360
QY 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLFSETVKLESLSKQES 420
Db 361 POPAPNLKIDSNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLFSETVKLESLSKQES 420
QY 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAPLAPITHTPERLKGKNSQIEYTEDVRIAQLADKYTTSDGYIFDHDHDIISD 540
Db 481 KEKLVDDLLAPLAPITHTPERLKGKNSQIEYTEDVRIAQLADKYTTSDGYIFDHDHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKIGILPSPDADVKANPTGDSAAAIY 600
Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKIGILPSPDADVKANPTGDSAAAIY 600
QY 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAFDDHTYKAPNGYTTLED 660
Db 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPKHQHYHNIKFAFDDHTYKAPNGYTTLED 660
QY 661 LPAITKYVVEHPDERPHSDGNGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
Db 661 LPAITKYVVEHPDERPHSDGNGNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
QY 721 QVETEKVBAQLKEAEVLLAKVTDSSLKANATETLAGLRNLTQIMDNNSIMAEAEKLLA 780
Db 721 QVETEKVBAQLKEAEVLLAKVTDSSLKANATETLAGLRNLTQIMDNNSIMAEAEKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796
RESULT 2
ABP54584
ID ABP54584 standard; protein; 796 AA.
XX
XX AC ABP54584;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP036 protein sequence SEQ ID NO:56.
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX Streptococcus pneumoniae.
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX WPI; 2002-479261/51.
DR N-PSDB; ABQ84819.
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.
XX Claim 11; Page 27; 70pp; English.
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcus pneumoniae infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX Sequence 796 AA;
SQ

Query Match 100.0%; Score 4163; DB 5; Length 796;
Best Local Similarity 100.0%; Pred. No. 9.4e-299;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

QY 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEIVNEVKGKGVIVKDGKYYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEIVNEVKGKGVIVKDGKYYVYLKDA 120

QY 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGVALARSQGYTTDGGYIFNADSIIDET 180
Db 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGVALARSQGYTTDGGYIFNADSIIDET 180

QY 181 GDVAVPHGDHYHVPKNEISASLAAAEAPLSGRGNLSRTYRRQNSDTSRTNWVPS 240
Db 181 GDVAVPHGDHYHVPKNEISASLAAAEAPLSGRGNLSRTYRRQNSDTSRTNWVPS 240

QY 241 VSNPGTNTNTSNNSTNSQASQNDIDSLLKQYKLPQSRHVESGLVDFDPAQITST 300
Db 241 VSNPGTNTNTSNNSTNSQASQNDIDSLLKQYKLPQSRHVESGLVDFDPAQITST 300

QY 301 ARGVAVPHGDHYHVPYQMSSELEERITARIIPLYRSNHWVPSRQPSQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHVPYQMSSELEERITARIIPLYRSNHWVPSRQPSQPTPEPSPG 360

QY 361 PQAPNLKIDSNSLSVQLVRKVGEGVVEKIGISRYVFAKOLPSETVKNLESKLKQES 420
Db 361 PQAPNLKIDSNSLSVQLVRKVGEGVVEKIGISRYVFAKOLPSETVKNLESKLKQES 420

QY 421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480
Db 421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480

QY 481 KEKLVDDLLAFLAPITHPERLGNPSQIETEDVEVRIQLADKYTTSDGYIFDEHDIISD 540
Db 481 KEKLVDDLLAFLAPITHPERLGNPSQIETEDVEVRIQLADKYTTSDGYIFDEHDIISD 540

QY 541 EGDVAVTPHMGSHWIGKDSLSDEKVAQAQYKEKGIILPPSPDADVKANPTGDSAAATY 600
Db 541 EGDVAVTPHMGSHWIGKDSLSDEKVAQAQYKEKGIILPPSPDADVKANPTGDSAAATY 600

QY 601 NRKVGKRIPLVRLPVWVHTVEVKNGLIIPKOHYHNIKEAFDDHTYKAPNGYTTLED 660
Db 601 NRKVGKRIPLVRLPVWVHTVEVKNGLIIPKOHYHNIKEAFDDHTYKAPNGYTTLED 660

QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKADEEPVEETPAEPEVP 720
Db 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKADEEPVEETPAEPEVP 720

QY 721 QVETEKVEAQLKEAEVLLAKVTDSSILKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780
Db 721 QVETEKVEAQLKEAEVLLAKVTDSSILKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780

QY 781 LLKGSNPFSSVSKEKIN 796
Db 781 LLKGSNPFSSVSKEKIN 796

RESULT 3
ADC45137
ID ADC45137 standard; protein; 796 AA.
XX
AC ADC45137;
XX
DT 18-DEC-2003 (first entry)
XX
DE S. pneumoniae antigenic protein SP036.
XX
KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
OS Streptococcus pneumoniae.
XX
PN US6573082-B1.
XX
PD 03-JUN-2003.
XX
PF 28-MAR-2000; 2000US-00536784.
XX
PR 31-OCT-1996; 96US-0029960P.
PR 30-OCT-1997; 97US-00961083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
DR WPI; 2003-764574/72.
DR N-PSDB; ADC45136.
XX
PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT useful for producing vaccines for prevention or attenuation of infection
PT by Streptococcus pneumoniae.
XX
PS Example 1; SEQ ID NO 56; 58pp; English.
XX
CC The invention relates to an isolated polynucleotide consisting of a
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
CC antigens. Also included are making a recombinant vector by inserting the
CC nucleic acid into a vector, an isolated polynucleotide consisting of at
CC least 50 or 100 contiguous nucleotides of the SP028 polynucleotide, and a
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
CC acids are useful as DNA vaccine against Streptococcus pneumoniae
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
CC antigen nucleic acids are useful as probes for use in diagnostic methods
CC for detecting S. pneumoniae gene expression. The present sequence
CC represents an S. pneumoniae antigenic protein.
XX
SQ Sequence 796 AA;

Query Match 100.0%; Score 4163; DB 7; Length 796;
Best Local Similarity 100.0%; Pred. No. 9.4e-299;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNKLKDEIDVNEVKGYYIKVDGKYVYLKDA 120
 Db 61 SHGDHYHYNGKVPYDAIISELLMKDPNKLKDEIDVNEVKGYYIKVDGKYVYLKDA 120
 QY 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 Db 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 QY 181 GDVIVPHGDHYHYIPKNELASASIAAFLSGRNLNSRTYRRQNSDNTSRNWVPS 240
 Db 181 GDVIVPHGDHYHYIPKNELASASIAAFLSGRNLNSRTYRRQNSDNTSRNWVPS 240
 QY 241 VSNPGTNTNNSNTNSQASQNDISLILKQYKPLPSORHVESDGLVFDPAQITSR 300
 Db 241 VSNPGTNTNNSNTNSQASQNDISLILKQYKPLPSORHVESDGLVFDPAQITSR 300
 QY 301 ARGVAVPHGDHYHETPYQMSSELEBRIARIIPLRVRSNHWPDSRPEQSPQTPPEPSG 360
 Db 301 ARGVAVPHGDHYHETPYQMSSELEBRIARIIPLRVRSNHWPDSRPEQSPQTPPEPSG 360
 QY 361 PQAPNLKIDNSSLSVQLVRKVGEGYVPEEKISRYVFAKDLPSRTVKNLESKLSQES 420
 Db 361 PQAPNLKIDNSSLSVQLVRKVGEGYVPEEKISRYVFAKDLPSRTVKNLESKLSQES 420
 QY 421 VSHTLTAKENVAPRQDFYDQYKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 Db 421 VSHTLTAKENVAPRQDFYDQYKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 QY 481 KEKLVDDLLAFAPITTHPERLKGPNQSIETEDVRIQADKYTTSDGYIFDEHDIISD 540
 Db 481 KEKLVDDLLAFAPITTHPERLKGPNQSIETEDVRIQADKYTTSDGYIFDEHDIISD 540
 QY 541 EGDVYTPMHGSHWIKGDSLSDKEKVAQAQVYKKGILPPSPDADVKANPTGDSAAIY 600
 Db 541 EGDVYTPMHGSHWIKGDSLSDKEKVAQAQVYKKGILPPSPDADVKANPTGDSAAIY 600
 QY 601 NRKVGKRIPLVRLPMVHTVEKGNLIIPKHQVHNIKPAWFDHDKYKAPNGYTLSD 660
 Db 601 NRKVGKRIPLVRLPMVHTVEKGNLIIPKHQVHNIKPAWFDHDKYKAPNGYTLSD 660
 QY 661 LFAATIKYVHEPDERPHSNDGWSAGHVLGKDHSDPNKFNKADPEEVEETPAPEVP 720
 Db 661 LFAATIKYVHEPDERPHSNDGWSAGHVLGKDHSDPNKFNKADPEEVEETPAPEVP 720
 QY 721 QVETEKVEAQLKEAEVLLAKVTSSKANATETIAGLRNMLTIQIMDNNSIMAEKLLA 780
 Db 721 QVETEKVEAQLKEAEVLLAKVTSSKANATETIAGLRNMLTIQIMDNNSIMAEKLLA 780
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 781 LLKGSNPSSVSKEKIN 796

RESULT 4
 ABU01598
 ID ABU01598 standard; protein; 802 AA.
 XX AC ABU01598;
 XX AC
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX DE
 DE S. pneumoniae type 4 strain protein from coding region #1174.
 XX DE
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX OS
 OS Streptococcus pneumoniae; type 4 strain.
 XX PN
 PN WO200277021-A2.
 XX XX
 XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB002163.
 XX 27-MAR-2001; 2001GB-00007658.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX06886.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 XX Claim 1; SEQ ID NO 2348; 56pp; English.
 PS
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX
 SQ

Query Match 100.0%; Score 4163; DB 6; Length 802;
 Best Local Similarity 99.9%; Pred. No. 9.5e-299;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
 Db 7 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 66
 QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNKLKDEIDVNEVKGYYIKVDGKYVYLKDA 120
 Db 67 SHGDHYHYNGKVPYDAIISELLMKDPNKLKDEIDVNEVKGYYIKVDGKYVYLKDA 126
 QY 121 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 Db 127 AHADNVTKEINRQKQEHSHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 186
 QY 181 GDVIVPHGDHYHYIPKNELASASIAAFLSGRNLNSRTYRRQNSDNTSRNWVPS 240
 Db 187 GDVIVPHGDHYHYIPKNELASASIAAFLSGRNLNSRTYRRQNSDNTSRNWVPS 246

QY 241 VSNPGTNTNTSNTNSQASQSDNDISLLKQYKPLSQRHVESDGLVFPQAITSRT 300
 Db 247 VSNPGTNTNTSNTNSQASQSDNDISLLKQYKPLSQRHVESDGLVFPQAITSRT 306
 QY 301 ARGVAVPHGDHVFIPYSQMSLEERIIARIIPLRVSRNHVWVDSRPEQSPQTPSPG 360
 Db 307 ARGVAVPHGDHVFIPYSQMSLEERIIARIIPLRVSRNHVWVDSRPEQSPQTPSPG 366
 QY 361 POPAPNLKIDSNSLSVQVVRVGVGVEEKGISRYVFAKDLPSRTVKNLESLSKQES 420
 Db 367 POPAPNLKIDSNSLSVQVVRVGVGVEEKGISRYVFAKDLPSRTVKNLESLSKQES 426
 QY 421 VSHTLTAKENVAPRDQEFYDYNLTLTAHKAFLXNKGNSDFQALDKLLERLNDESTN 480
 Db 427 VSHTLTAKENVAPRDQEFYDYNLTLTAHKAFLXNKGNSDFQALDKLLERLNDESTN 486
 QY 481 KEKLVDDLLAFAPITTHPERLGKPSQIEYTEDVRIIAQLADKYTTSQYIPDEHDIISD 540
 Db 487 KEKLVDDLLAFAPITTHPERLGKPSQIEYTEDVRIIAQLADKYTTSQYIPDEHDIISD 546
 QY 541 EGDYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKNPTGSAAIY 600
 Db 547 EGDYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKNPTGSAAIY 606
 QY 601 NRKVGKRIPLVRPLPYVVEHTVEVKNGLIIPHKDHYHNIKFAPDDHTYKAPNGYTTLED 660
 Db 607 NRKVGKRIPLVRPLPYVVEHTVEVKNGLIIPHKDHYHNIKFAPDDHTYKAPNGYTTLED 666
 QY 661 LFATIKYVVEHDPDRSHNDGWSNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
 Db 667 LFATIKYVVEHDPDRSHNDGWSNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 726
 QY 721 QVTEKVEAQKAEVLLAKVTDSSKANATETIAGLRNNLTLOIMDNNSIMAEAEKLLA 780
 Db 727 QVTEKVEAQKAEVLLAKVTDSSKANATETIAGLRNNLTLOIMDNNSIMAEAEKLLA 786
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 787 LLKGSNPSSVSKEKIN 802

RESULT 5

AAB01468
 ID AAB01468 standard; protein; 819 AA.
 AC AAB01468;
 XX
 20-OCT-2000 (first entry)
 DE Recombinant variant of Sp36 (Sp36A) of *S. pneumoniae*.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia.
 XX
 OS Streptococcus pneumoniae.

Key Location/Qualifiers
 FT Region 63..68
 FT Region /label= Histidine triad residue
 FT Region 118..145
 FT Region /label= Coiled coil region
 FT Region 189..194
 FT Region /label= Histidine triad residue
 FT Region 309..314
 FT Region /label= Histidine triad residue
 FT Region 406..434
 FT Region /label= Coiled coil region
 FT Region 462..493
 FT Region /label= Coiled coil region
 FT Region 550..555
 FT Region /label= Histidine triad residue

FT Region 634..639
 FT /label= Histidine triad residue
 FT 724..751
 FT /label= Coiled coil region
 XX
 PN WO200037105-A2.
 PD 29-JUN-2000.
 XX
 PF 21-DEC-1999; 99WO-US030390.
 XX
 PR 21-DEC-1998; 98US-0113048P.
 XX
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Johnson LS, Koenig S, Adamou JE;
 XX
 DR WPI; 2000-452129/39.
 DR N-PSDB; AAA47604.
 XX
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX
 PS Claim 1; Page 61-64; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of *S. pneumoniae* in order to utilize such polypeptides in vaccines to
 CC protect against a wide variety of *S. pneumoniae*. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX
 SQ Sequence 819 AA;

Query Match 100.0%; Score 4163; DB 3; Length 819;
 Best Local Similarity 99.9%; Pred. No. 9.8e-299;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
 Db 21 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
 QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKDA 120
 Db 81 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKDA 140
 QY 121 AHADNVRTKEEINRQKQHSQHRGGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 Db 141 AHADNVRTKEEINRQKQHSQHRGGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 200
 QY 181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRTNWVPS 240
 Db 201 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRTNWVPS 260
 QY 241 VSNPGTNTNTSNTNSQASQSDNDISLLKQYKPLSQRHVESDGLVFPQAITSRT 300
 Db 261 VSNPGTNTNTSNTNSQASQSDNDISLLKQYKPLSQRHVESDGLVFPQAITSRT 320
 QY 301 ARGVAVPHGDHVFIPYSQMSLEERIIARIIPLRVSRNHVWVDSRPEQSPQTPSPG 360
 Db 321 ARGVAVPHGDHVFIPYSQMSLEERIIARIIPLRVSRNHVWVDSRPEQSPQTPSPG 380
 QY 361 POPAPNLKIDSNSLSVQVVRVGVGVEEKGISRYVFAKDLPSRTVKNLESLSKQES 420
 Db 381 POPAPNLKIDSNSLSVQVVRVGVGVEEKGISRYVFAKDLPSRTVKNLESLSKQES 440

RESULT 7	
AAB12727	
ID	AAB12727 standard; protein; 821 AA.
XX	
AC	AAB12727;
XX	
DT	21-NOV-2000 (first entry)
XX	

CC	antigen, from the present invention	
XX		
QQ	Sequence 821 AA:	
	Query Match	77.3%; Score 3218; DB 3; Length 821;
	Best Local Similarity	75.2%; Pred. No. 8.2e-229;
	Matches 616; Conservative	65; Mismatches 104; Indels 34; Gaps 3;
QY	1 SYELGLVQARTKNNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGYVT	60
DB	2 AYELGLHQACTKNNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGYVT	61
QY	61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDIIVNEKGGYIVKYDGKVVYVLKDA	120
DB	62 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDIIVNEKGGYIVKYDGKVVYVLKDA	121
QY	121 AHADNVTKEINRQKQEHSGHEGGTPRNDGAVALLARSOCRYTDDGYIFNADIIEDT	180
DB	122 AHADNVTKEINRQKQEHSGHEGGTSANDGAVAFARSOCRYTDDGYIFNADIIEDT	181
QY	181 GDVIVPHGDHYHYPKNELASASELAFLSGRGNLSNRYRRNSONTGRTNWVPS	240
DB	182 GDVIVPHGDHYHYPKNELASASELAFLSGRENLSNLTYYRRNSONTGRTNWVPS	241
QY	241 VSNPGTTNTNNTSNNTNSOASQNDISLLKQLYKPLSQRVHVESGLVDFDPAQITSR	300
DB	242 VSNPGTTNTNNTSNNTNSOASQNDISLLKQLYKPLSQRVHVESGLVDFDPAQITSR	301
QY	301 ARGVAVPHGDHYHFTPYQSOMSELEERTARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG	360
DB	302 ARGVAVPHGNHYHFTPYEQMSELEKRIARIIPLYRSNHWVPDSRPPQSPQPTPEPSPS	361
QY	361 POPAPNLK-IDSN---SSLVSQLVKRVKGGYVFEKGISRVVFAKDLPSETVKMLSKLS	416
DB	362 POPAPNPQAPNSNPIDEKLVKEAVKVGQGVFEENGVSRYIPAKNLSAETAAGIDSKLA	421
QY	417 KOESVSHLTAKKNVAPRDOEFTDKAYNLTAEHALFYKNGKNSDFOALDKLLERLND	476
DB	422 KOESLSHLGAKKTDLPSDREFYNKAYDILLARIHQDLDNKRQGVDFEALDNLLERLKD	481
QY	477 ESTNKEKLVDDLAFAPLTHPERLGKPNISOIEVTEDEVRIAQADKYTTSDGYIFDHD	536
DB	482 VSSDKVKLVDDLAFAPLTHPERLGKPNQAQITTYDDBIQAVALAGKTYTDEGYIFDPRD	541
QY	537 IISDSGDVAVTPHMGHSHWIKGDSLSDEKVAQAAYTKBKGLPPSPDADVKNPTGQSA	596
DB	542 IYSDSGDAVTPHMTSHWIKKDSLSAEARAQAAYAKEKGLTPSTDHQDSGNTAEKA	601
QY	597 AAIYNRVKGEKRIPLVRUPYVYVHTVFNKGNLLIIPHKDHYHNIKFAFDHHTYKPNXY	656
DB	602 EAIYNRVKAARKVPLDRMPYNLQYTVFNKGNLLIIPHYDHYHNIKFEWDFEGLYEA	661
QY	657 TLEDLIPATYKYVVEHPDHPHSDNGWGNASHEVLGKSDHSDPNKNFKAD-----	706
DB	662 TLEDLIAVTKYVVEHPNERPHSDNGFGNASDHWQNKNGQADTNQTEKPEEKPTKEPE	721
QY	707 -----EETPETAEPVPPQVETEKVEAQLKEAEVLLAKVYDSSL	746
DB	722 EETPREEKQSEKPSKPTETPEEESPESEEPQVETEKVEEKLREABDILGKIQDPTI	781
QY	747 KANATETLAGLRNLTQIIMDNNISMAEAKLLALLKGS	795
DB	782 KSNAKETVNGKKNLTFGTDNNITMAEAKLLALLKES	820

RESULT 7	
AAB12727	
ID	AAB12727 standard; protein; 821 AA.
XX	
AC	AAB12727;
XX	
DT	21-NOV-2000 (first entry)
XX	

Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA001218.

23-DEC-1998; 98US-0113800P.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N; WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteraemia and/or pneumonia.

Claim 18; Fig 25; 106pp; English.

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-11M protein antigen

Sequence 821 AA;

Query Match 77.3%; Score 3218; DB 3; Length 821;

Best Local Similarity 75.2%; Pred. No. 8.2e-229;

Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

1 SYELGVLQARTVKNRVSVDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVT 60

2 AYELGLHQAQVKNRVSVDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVT 61

61 SHGDHYHYNGKVPYDAIIEELLMDPNYKLDXEDIVNEVKGGYIVKDGYYVYLKDA 120

62 SHGDHYHYNGKVPYDAIIEELLMDPNYKLDXEDIVNEVKGGYIVKDGYYVYLKDA 121

121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTDDGYIFNASDIIEDT 180

122 AHADNVRTKEEINRQKQHSQHREGGTSNDGAVAFARSGRYTDDGYIFNASDIIEDT 181

181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240

182 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 241

241 VSNPGTTNTNTNNTNSQASQSDNIDSLKLYKLPISQRHVSDGLVFPDPAQITSRT 300

242 VSNPGTTNTNTNNTNSQASQSDNIDSLKLYKLPISQRHVSDGLVFPDPAQITSRT 301

301 ARGVAVPHGDHYHYIPYQMSQSELEERARIIPLYRVSNEHWVDSRPEQSPQTPPEPSG 360

302 ARGVAVPHGNHYHYIPYQMSQSELEERARIIPLYRVSNEHWVDSRPEQSPQTPPEPSG 361

361 POPAPNLK-IDSN---SSIVSQLVKRGYVFEKGISRYVFAKDLPSQETVKNLESKLS 416

362 POPAPNPQAPNPIDEKLVKAEAVKRGYVFEENGVSRYIPAKNLSAETAGDISKLA 421

417 KOESVSHITLAKENVAPRQDFYQKAYNLLTAHKAFLXNKGNSDFQALDKLLERLND 476

Db 422 KOESLSHLGAKKTDLPSSDRFYNNKAYDOLLARIHQDLNDKGRQVDFEALDNLRLKXD 481

QY 477 ESTNKEKLVDDLLAFAPITHPERLCKPNSQIETDEVRIAQLADKYTTSQGYIEDEHD 536

Db 482 VSSDKVLVDLLAFAPIRHPERLCKPNSQIETDEVRIAQLADKYTTSQGYIEDEHD 541

QY 537 IISDEGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 596

Db 542 IISDEGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKKGLPSPDADVKANPTGDSA 601

QY 597 AAIYNRVKGEKIPLVRLPYVVEHTVEKNGNLIIPKDHYNHNIKPAFDDHHTYKAPNGY 656

Db 602 EAIYNRVKAAKVPIDRMPYQLQYTVKNGSLIIPHYDHYHNIKFEWDFEGLYKAPNGY 661

QY 657 TLEDLFATIKYVVEHPDHPHSDNGWNASHLGKDKHSDPDNKNFKAD----- 706

Db 662 TLEDLFATIKYVVEHPDHPHSDNGWNASHLGKDKHSDPDNKNFKAD----- 706

QY 707 -----BEPVEETPAPEVPOVETEKVEAQLKEAEVLIAKVTDSLSL 746

Db 722 BETPREKQSEKPEPKPTPEPESESESESESESESESESESESESESESESESESESE 781

QY 747 KANATETLAGLRNLTQLIMDNNSITMAEAEKLLALLKGS 785

Db 782 KSNAKEHLTGLKNNLLFGTQDNNTIMAEAEKLLALLKGS 820

RESULT 8

AAU84026

ID AAU84026 standard; peptide; 821 AA.

XX AAU84026;

AC AAU84026;

DT 08-MAY-2002 (first entry)

XX Truncated variant of S. pneumoniae BVH-11, BVH-11M.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; mutein.

OS Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

DN 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA000908.

XX 20-JUN-2000; 2000US-0212683P.

PR (SHIR-) SHIRE BIOCHEM INC.

PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

PI WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia.

XX Example 1; Page: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (iii) encoding (i) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (iii) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention

Sequence 821 AA;

Query Match	77.3%;	Score 3218;	DB 5;	Length 821;
Best Local Similarity	75.2%;	Pred. No. 8.2e-229;		
Matches 616;	Conservative 65;	Mismatches 104;	Indels 34;	Gaps 3;
QY	1	SYELGLYQARTVKNENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT	60	
Db	2	AYELGLHQAQTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT	61	
QY	61	SHGDHYHYNGKVPYDAIISELLMKDPNYKLDIEDIVNEKGVYKVDGKYVYVLKDA	120	
Db	62	SHGDHYHYNGKVPYDAIISELLMKDPNYKLDSDIVNEIKGGVYKVGKYYVYVLKDA	121	
QY	121	AHADNVRTKEENRQKQHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNADIIEDT	180	
Db	122	AHADNVRTKEENRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNADIIEDT	181	
QY	181	GDAYIVPHGDHYHYIIPKNELSAELAAEFISGRGNLSNRTYRQNSDNTSRINWVPS	240	
Db	182	GDAYIVPHGDHYHYIIPKNELSAELAAEFISGRGNLSNRTYRQNSDNTSRINWVPS	241	
QY	241	VSNPGTTNTNNSNTNSQASOSNDISLLKQYKPLSORHVESDGLVFPQAQITST	300	
Db	242	VSNPGTTNTNNSNTNSQASOSNDISLLKQYKPLSORHVESDGLVFPQAQITST	301	
QY	301	ARGVAVPHGDHYHYIIPYQMSSELEERARIIPLYRSNHWPDSPRPEQSPQPTPEPSPG	360	
Db	302	ARGVAVPHGNHYHYIIPYQMSSELEERARIIPLYRSNHWPDSPRPEQSPQPTPEPSPS	361	
QY	361	POAPNLK-IDSN---SSLVQLVRKVGEGYFEEKGISRYVFAKPLPSETVXNLSKLS	416	
Db	362	POAPNPQPAPSNPIDEKLKVEARVKVGDGVFEENGVSRYIIPAKNLSAETAAGIDSKLA	421	
QY	417	KQESVSHLTAKENVAPRDOEFVDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND	476	
Db	422	KQESLSHGKGAKTDLDPSSDREFYKAYDILLARIHQDLNKGQVDFEALDNLLEKLD	481	
QY	477	ESTNKEKLVDDLLAFLAPIHPERLPGKPSQIETEDEVRIAGLADKYSYDGIYFDEHD	536	
Db	482	VSDKVKLVDDLLAFLAPIRHPERLGPNAQITVTDDEIQVAKLAGKYTTEDGYIFDPRD	541	
QY	537	IISDEGAYVTPHNGHSHWTKGKSLDKKVAQAQYTKKGLPPSPDADVKANPTGDSA	596	
Db	542	ITSDEGAYVTPHNTSHWTKGKSLSEAEARAAQYAKKGLTPSPDTHQDSNGNTEAKGA	601	
QY	597	AATNVRKGEKRIPLVLPYVMTHTVGVGNLLIIPKHQVHNKIFKFAWFDHTYKAPNGY	656	
Db	602	EATNVRKAARKVPLDRMPYLNQITVEVNGSLIIPHYDHYHNKIFKFWFDEGLYEAPKG	661	
QY	657	TLEDLFATIKYVVEHPDHPHNSDGNWGNASHVILGKXHSDDPNKFKAD-----	706	
Db	662	TLEDLLATKYVVEHEPERHPSDNGFNGASDHVQKNKGQADTNTQTEKPSPEKQTEKPE	721	
QY	707	-----REPVEETPAPEVQVETEKVEAQLKAEVLAKVTDSSL	746	
Db	722	BETPREEKPSQSEKPSKPTPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEE	781	
QY	747	KANATETLAGLRNNLTQIMDNNSIMAEKILLALLKGS	785	

Db 782 KSNAKETTLTGKNNLLFGTQDNNTIMAEKILLALLKES 820

RESULT 9

ABMI18807

ID ABMI18807 standard; protein; 821 AA.

XX AC ABMI18807;

XX 13-OCT-2003 (first entry)

XX S. pneumoniae variant protein BVH-11M.

XX antibacterial; antiinflammatory; auditory; vaccine; meningitis;

XX streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;

XX mutant; mutein; New 43; BVH-11; BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO2003054007-A2.

XX 03-JUL-2003.

XX 20-DEC-2002; 2002WO-CA002006.

XX 20-DEC-2001; 2001US-0341252P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;

XX Labbe S;

XX WPI; 2003-569224/53.

XX New isolated polypeptides of Streptococcus pneumoniae, useful for

diagnosing, preventing or treating streptococcal infection, meningitis,

otitis media, bacteraemia or pneumonia infection.

XX Example 1; SEQ ID NO 19; 79pp; English.

XX The invention relates to a novel isolated polypeptide of Streptococcus

pneumoniae. A polypeptide of the invention has antibacterial,

antiinflammatory, and auditory activity, and is used as a vaccine. The

polypeptide or pharmaceutical composition is useful for the prophylactic

or therapeutic treatment of streptococcal infection, meningitis, otitis

media, bacteraemia or pneumonia infection. The kit is useful for

detecting or diagnosing streptococcal infection. The pharmaceutical

composition is useful as a vaccine. The polynucleotides are useful in

designing DNA probes for detecting circulating Streptococcus in a

biological sample. The present sequence is used in the exemplification of

the invention. Note: The sequence data for this patent is not fully

represented in the printed specification, but is based on sequence

information supplied by the European Patent Office

XX Sequence 821 AA;

Query Match

Best Local Similarity 77.3%; Score 3218; DB 7; Length 821;

Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKNENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60

2 AYELGLHQAQTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 61

61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLDIEDIVNEKGVYKVDGKYVYVLKDA 120

62 SHGDHYHYNGKVPYDAIISELLMKDPNYKLDSDIVNEIKGGVYKVGKYYVYVLKDA 121

121 AHADNVRTKEENRQKQHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNADIIEDT 180

122 AHADNVRTKEENRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNADIIEDT 181

181 GDYIVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 240
 182 GDYIVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 241
 241 VSNPGTNTNTSNTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
 242 VSNPGTNTNTSNTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 301
 301 ARGVAVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 360
 302 ARGVAVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 361
 361 PQAPNLK-IDSN---SSLVQLVKRKGEGYVFEKGSRYVFAKDLPSVTKNLESKLS 416
 362 PQAPNPQAPNSPIDEKLVKAEVRKVGDFVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
 417 KQESVSHLTAKENVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 476
 422 KQESLSHLGAKKTDLPSDREFYNKAYDILLARIHQDLNDKGRQVDFEALDNLRLKD 481
 477 ESTNKEKLVDDLLAFAPITHPERLKPNSQIETEDVRIQAQADKYTSDGYIFDEHD 536
 482 VSSDKVKLVDDLLAFAPITHPERLKPNSQIETEDVRIQAQADKYTSDGYIFDEHD 541
 537 IISDEGDYVTPHMGSHWIGKDSLSDEKVAQAAYTKKGLPSPDADVKANPTGDSA 596
 542 IISDEGDYVTPHMGSHWIGKDSLSDEKVAQAAYTKKGLPSPDADVKANPTGDSA 601
 597 AAIYNRVKGKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWDDHTYKAPNGY 656
 602 EAIYNRVKAACKVPLDRMPYNLQVTEVKNGLIIPHYDHYHNIKFEWDEGLYEAPKGY 661
 657 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSDPNKNPKAD----- 706
 662 TLEDLATVYVVEHPDERPHSDNGWGNASEHVLGKDHSDPNKNPKAD----- 721
 707 -----EPPVEETPABPEVPOVETEKVEAQLEAEVLLAKVITDSSL 746
 722 EETPREKQSEKPEPKTEEPSEESPESEBPQVETEKVEKLEAEVLLAKVITDSSL 781
 747 KANATETAGLRNLTQIMDNNSIMAEKLLALLKGS 785
 782 KSNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKES 820

RESULT 10
 AAB12716
 ID AAB12716 standard; protein; 840 AA.
 AC AAB12716;
 AC AAB12716;
 DT 21-NOV-2000 (first entry)
 XX
 XX
 DE Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX Streptococcus pneumoniae.
 OS
 XX WO200039299-A2.
 PN
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.
 DR N-PSDB; AAA65731.
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 XX Claim 18; Fig 4; 106pp; English.
 PS
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence represents the S. pneumoniae BVH-11
 CC protein antigen
 XX
 XX Sequence 840 AA;
 SQ
 Query Match 77.3%; Score 3218; DB 3; Length 840;
 Best Local Similarity 75.2%; Pred. No. 8.5e-229;
 Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
 QY 1 SYELGLYQARTVKNRVSVIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
 DB 21 AYLGLHQAQTVKNRVSVIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
 QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKEDIIVNEVKGYYIVKDGKYYVYLKDA 120
 DB 81 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKEDIIVNEVKGYYIVKDGKYYVYLKDA 140
 QY 121 AHADNVRTKEEINRQKQESHQHREGGTPRNDGVALARSOGRYTTDDGYIFNASDIIEDT 180
 DB 141 AHADNVRTKEEINRQKQESHQHREGGTSNDGAVAFARSOGRYTTDDGYIFNASDIIEDT 200
 QY 181 GDYIVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 240
 DB 201 GDYIVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 260
 QY 241 VSNPGTNTNTSNTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
 DB 261 VSNPGTNTNTSNTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 320
 QY 301 ARGVAVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 360
 DB 321 ARGVAVPHGDHYHYPKNEISASLSELAFAFLSGRNLNSRTYRQNSDNTSRNTWVPS 380
 QY 361 PQAPNLK-IDSN---SSLVQLVKRKGEGYVFEKGSRYVFAKDLPSVTKNLESKLS 416
 DB 381 PQAPNPQAPNSPIDEKLVKAEVRKVGDFVFEENGVSRYIPAKNLSAETAAGIDSKLA 440
 QY 417 KQESVSHLTAKENVAPRQDFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLRLND 476
 DB 441 KQESLSHLGAKKTDLPSDREFYNKAYDILLARIHQDLNDKGRQVDFEALDNLRLKD 500
 QY 477 ESTNKEKLVDDLLAFAPITHPERLKPNSQIETEDVRIQAQADKYTSDGYIFDEHD 536
 DB 501 VSSDKVKLVDDLLAFAPITHPERLKPNSQIETEDVRIQAQADKYTSDGYIFDEHD 560
 QY 537 IISDEGDYVTPHMGSHWIGKDSLSDEKVAQAAYTKKGLPSPDADVKANPTGDSA 596
 DB 561 IISDEGDYVTPHMGSHWIGKDSLSDEKVAQAAYTKKGLPSPDADVKANPTGDSA 620
 QY 597 AAIYNRVKGKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWDDHTYKAPNGY 656
 DB 621 EAIYNRVKAACKVPLDRMPYNLQVTEVKNGLIIPHYDHYHNIKFEWDEGLYEAPKGY 680
 QY 657 TLEDLFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSDPNKNPKAD----- 706
 DB 681 TLEDLATVYVVEHPDERPHSDNGWGNASEHVLGKDHSDPNKNPKAD----- 740

QY 707 -----EEPVETPAEVPQVETEKVQAOLKEAEVLAKVTDSSL 746
 Db 741 EETPREKQSEKPEPKTEPEEPESEEPQVETEKVQREAEADLLGKIQDPII 800
 QY 747 KANATETLAGLRNNLTQIMDNNSIMAEAEKLLALLKGS 785
 Db 801 KSNAKETITGLKNLLFTQDNTIMAEAEKLLALLKES 839

RESULT 11
 AAU75933
 ID AAU75933 standard; protein; 840 AA.
 AC AAU75933;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-11 protein.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
 KW streptococcal bacterial infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA000908.
 XX
 PR 20-JUN-2000; 2000US-0212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI
 DR WPI; 2002-122272/16.
 DR N-PSDB; ABK15103.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
 PT polypeptides, useful as vaccine components for treating or preventing
 PT streptococcal infections such as otitis media, meningitis, and
 PT bacteraemia.
 XX
 PS Example 1; Fig 7; 113pp; English.
 CC
 CC The invention describes an isolated polypeptide (I) with 70-90% identity
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This is the amino acid sequence of
 CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic
 CC peptides described in the method of the invention
 XX
 SQ Sequence 840 AA;
 XX
 Query Match 77.3%; Score 3218; DB 5; Length 840;
 Best Local Similarity 75.2%; Pred. No. 8.5e-229;
 Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKNENRVSYIDGKQATOKTENLTPEDEVSKREGINARQIVIKITDQGYVT 60
 Db 21 AYELGLHQAQTVKNENRVSYIDGKQATOKTENLTPEDEVSKREGINARQIVIKITDQGYVT 80
 QY 61 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEKGGYVILKDGKYYVYLKDA 120
 Db 81 SHGDHYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEKGGYVILKVGKYYVYLKDA 140
 QY 121 AHADNVRTKEINRQKOEHSQHREGGTPRNDGVALARSQGRYTTDDGYIFNASSDIIBDT 180
 Db 141 AHADNVRTKEINRQKOEHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASSDIIBDT 200
 QY 181 GDAYIVPHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNRTYRRQNSDNTSRNWVPS 240
 Db 201 GDAYIVPHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNRTYRRQNSDNTSRNWVPS 260
 QY 241 VSNFGTTNTNNTSNNTSNQASQNDIDSLKQLYKPLSORHVESDGLVDFDPAQITSRT 300
 Db 261 VSNFGTTNTNNTSNNTSNQASQNDIDSLKQLYKPLSORHVESDGLIFDPAQITSRT 320
 QY 301 ARGVAVPHGDHYHYIPYQMSSELEBERIARIIPLYRSNHWVPDSRPEQSPQTPPEPSG 360
 Db 321 ARGVAVPHGDHYHYIPYQMSSELEBERIARIIPLYRSNHWVPDSRPEQSPQTPPEPSG 380
 QY 361 POPAPNLK-IDSN---SSLSQLVKVGEGYVPEKIGISRYVFAKDLSEITVKNLESKL 416
 Db 381 POPAPNPQAPSNPIDKLVKEAVRKVGDDGVFENGVSRYIPAKNLSAETAAGIDSKLA 440
 QY 417 KOESVSHLTAKENVAPRDOBFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLND 476
 Db 441 KOESLSHKLGAARKTDLPSDSREFYNKAYDLARITHQDLDNKGQVDFEALDNLLEKLD 500
 QY 477 ESTNKEKLVDDLLAFAPITHEPERLGKPNSEIYETDEVRIAQADKYTTSQGYIFDEHD 536
 Db 501 VSSDKVLVDLLAFAPIRHPERLGKPNQAQITTYDDEIQVAKLAGKYTTEDGYIFDPRD 560
 QY 537 TISDEGDAYVTPHMGSHWIKDLSLSEKVAQAAYTKEKGLLPSPADVKANPTGDSA 596
 Db 561 ITSDEGDAYVTPHMTSHWIKDLSLSEAEARAAQAYAKEKGLTPPSTDHQDSGNTAKGA 620
 QY 597 RAIYNRVKGEKIPILVRLPYVVEHTVEVKNGLNLIIPHDKDHVHNIFKAFDDHRTYKAPNGY 656
 Db 621 EAIYNRVKAAKVPDLRMPYNLQYIVVEVNGSLIIPHVDHYHNIFKFEWDEGLYEAPKGY 680
 QY 657 TLEDLFATIKYVVEHDERPHSDNGWGNASERHLKGDHSEDPNKNFKAD----- 706
 Db 681 TLEDLLAIVKYVVEHPNERPHSDNGFNGASDHVQRNKGQADNTQTEKPESEKPKQTEKPE 740
 QY 707 -----EEPVETPAEVPQVETEKVQAOLKEAEVLAKVTDSSL 746
 Db 741 EETPREKQSEKPEPKTEPEEPESEEPQVETEKVQREAEADLLGKIQDPII 800
 QY 747 KANATETLAGLRNNLTQIMDNNSIMAEAEKLLALLKGS 785
 Db 801 KSNAKETITGLKNLLFTQDNTIMAEAEKLLALLKES 839

RESULT 12
 ABM18797
 ID ABM18797 standard; protein; 840 AA.
 XX
 AC ABM18797;
 XX
 DT 13-OCT-2003 (first entry)
 XX
 DE S. pneumoniae BVH-11 polypeptide SEQ ID NO: 8.
 XX
 KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
 KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3.
 OS Streptococcus pneumoniae.
 XX
 PN WO2003054007-A2.

[illegible]

Matches	556;	Conservative	54;	Mismatches	75;	Indels	4;	Gaps	2;
XX	PA	(SHIR-) SHIRE BIOCHEM INC.							
XX	PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;							
XX	DR	WPI, 2002-122272/16.							
XX	PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing							
XX	PT	polypeptides, useful as vaccine components for treating or preventing							
XX	PT	streptococcal infections such as otitis media, meningitis, and							
XX	PT	bacteremia.							
XX	PS	Example 1; Page; 113pp; English.							
XX	CC	The invention describes an isolated polypeptide (I) with 70-90% identity							
XX	CC	to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or							
XX	CC	BVH-11, or chimeric sequences derived from them. A vaccine (II)							
XX	CC	comprising (I) is useful for therapeutic or prophylactic treatment of							
XX	CC	meningitis, otitis media, bacteraemia or pneumonia infection in an							
XX	CC	individual susceptible to these disorders. (II) is also useful for							
XX	CC	therapeutic or prophylactic treatment of any streptococcal bacterial							
XX	CC	infection (e.g., caused by Streptococcus pneumoniae, group A							
XX	CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such							
XX	CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or							
XX	CC	Staphylococcus aureus) in an individual susceptible to the infection. A							
XX	CC	polynucleotide (III) encoding (I) is useful in DNA immunisation							
XX	CC	techniques. The Streptococcus polypeptides are useful in a diagnostic							
XX	CC	test for S. pneumoniae infection. (III) is useful for designing DNA							
XX	CC	probes for use in detecting the presence of Streptococcus in a biological							
XX	CC	sample suspected of containing the bacteria. The DNA probes may also be							
XX	CC	used for detecting circulating S. pneumonia nucleic acid in a sample for							
XX	CC	diagnosing streptococcal infections. This sequence represents a truncate							
XX	CC	of a Streptococcus pneumoniae gene used to obtain antigenic peptides, as							
XX	CC	described in the method of the invention. Note: This sequence does not							
XX	CC	appear in the specification but has been created according to information							
XX	CC	given in the invention							
XX	SQ	Sequence 690 AA;							
		Query Match 71.11%; Score 2961; DB 5; Length 690;							
		Best Local Similarity 80.7%; Pred. No. 6.4e-210; Indels 4; Gaps 2;							
		Matches 556; Conservative 54; Mismatches 75;							
Qy	1	SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60							
Db	2	AYELGLHQAOQTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 61							
Qy	61	SHGDHYHYNGKVPYDAIISELLMKDPNYKQDSDIVNEIKGGYVKGKYYVYLKDA 120							
Db	62	SHGDHYHYNGKVPYDAIISELLMKDPNYKQDSDIVNEIKGGYVKGKYYVYLKDA 121							
Qy	121	AHADNVRTKEEINRQKQHSOHRREGTTPNDGAVALARSQGYTTDDGYIFNASDIIEDT 180							
Db	122	AHADNVRTKEEINRQKQHSOHRREGTTPNDGAVALARSQGYTTDDGYIFNASDIIEDT 181							
Qy	181	GDYIVPHGDHYHYIPKNELSASELAFAFLSGRGLNSRTYRQNSDNTSRTNWVPS 240							
Db	182	GDYIVPHGDHYHYIPKNELSASELAFAFLSGRGLNSRTYRQNSDNTSRTNWVPS 241							
Qy	241	VSNPGTTNTNTSNNTSNQASQSDNDISLLKQLYKPLSQRHVESDGLVDPQAITSRT 300							
Db	242	VSNPGTTNTNTSNNTSNQASQSDNDISLLKQLYKPLSQRHVESDGLVDPQAITSRT 301							
Qy	301	ARGVAVPHGDHYHYIPYQMSLELEBRIARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 360							
Db	302	ARGVAVPHGNHYHFIPIYQMSLELEKRIARIIPLYRSNHWVDSRPEEPSPQPTPEPSPS 361							
Qy	361	POAPNLK-IDSN---SSIVSQLVKRVGEGYVPEEKGISRYVFAKDLPSETVKNLESKLS 416							
Db	362	POAPNPQAPSNPIDKLVKAVRKVGQYVFEENGYSRYPKAKLSAETAAGIDSKLA 421							
Qy	417	KQESVSHTLTAKKENVAPRQDFYDKAVNLLTEAHKALFXNKGNSDFQALDKLLERLND 476							
Db	422	KQESLSHLKCAKTDLPSSDREFYNKAYDILLARIHQDLDNKGROVDPEALDNLLERLKD 481							
Qy	477	ESTNKGKLVDDLAFIAPITHPERLCKPNSQIETYEDEVRAQALADKYTTSDGYIFEBHD 536							
Db	482	VSSDKVKLVDDILAFIAPIRHPERLCKPNAQIYTTDDEIQVAKLAGKYTTEDGYIFDPRD 541							
Qy	537	IISDEGDYVTPHMGSHWIGKDSLSDEKVAQAQYTKKGLILPPSPADVKNAPTGDGA 596							
Db	542	ITSDEGDYVTPHMTSHHWIKDLSSEAEAAAQAYAKEKGLTPSTPDH							

[illegible]

Db 422 KQESLSHKLGAKTDLPSDDREFYKAYDILLARHQDLNDKGRQVDFEALDNLRLERKD 481
 QY 477 ESTNKEKLVDDLLAFAPITHTPERLGKPSQIEYTEDVIRIAQLADKYTTSDGYIFDEHD 536
 Db 482 VSDKVKLVDDLLAFAPITHTPERLGKPSQIEYTEDVIRIAQLADKYTTSDGYIFDEHD 541
 QY 537 IISDEGDAYVTPHMGSHWIGKSDLSDEKVAQAAYTKEKGLPSPDADVKANPTGDSA 596
 Db 542 ITSDEGDAYVTPHMGSHWIKKDSLSAEABRAAAQAYAKEKGLTPSTDHDSGNTAKGA 601
 QY 597 RAIYNRVKGEKPIPLVRLPYMVEHTVEKNGNLIIPHKDHYHNKFAWDDHTYKAPNGY 656
 Db 602 EAIYNRVKAAKVPDLRMPYNLQYTVVEKNGSLIIPHYDHYHNKFEWDEGLYEAPKY 661
 QY 657 TLEDLFATIKYVVEHDPDERPHSDNGWGNA 685
 Db 662 TLEDLLATVKYVVEHPNRPHERPHSDNGFGNA 690

RESULT 15
 ID ABM18826 standard; protein; 690 AA.
 AC ABM18826;
 XX
 DT 13-OCT-2003 (first entry)
 DE S. pneumoniae variant protein NEW16.
 KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;
 KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;
 KW mutant; mteuin; New 43; BVH-11; BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 OS
 PN WO2003054007-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 20-DEC-2002; 2002WO-CA002006.
 XX
 PR 20-DEC-2001; 2001US-0341252P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;
 PI Labbe S;
 XX
 DR WPI; 2003-569224/53.
 XX
 PT New isolated polypeptides of Streptococcus pneumoniae, useful for
 PT diagnosing, preventing or treating streptococcal infection, meningitis,
 PT otitis media, bacteraemia or pneumonia infection.
 PS
 PS Example 1; SEQ ID NO 38; 79pp; English.
 XX
 CC The invention relates to a novel isolated polypeptide of Streptococcus
 CC pneumoniae. A polypeptide of the invention has antibacterial,
 CC antiinflammatory, and auditory activity, and is used as a vaccine. The
 CC polypeptide or pharmaceutical composition is useful for the prophylactic
 CC or therapeutic treatment of streptococcal infection, meningitis, otitis
 CC media, bacteraemia or pneumonia infection. The kit is useful for
 CC detecting or diagnosing streptococcal infection. The pharmaceutical
 CC composition is useful as a vaccine. The polynucleotides are useful in
 CC designing DNA probes for detecting circulating Streptococcus in a
 CC biological sample. The present sequence is used in the exemplification of
 CC the invention. Note: The sequence data for this patent is not fully
 CC represented in the printed specification, but is based on sequence
 CC information supplied by the European Patent Office
 XX
 SQ Sequence 690 AA;

Query Match 71.1%; Score 2961; DB 7; Length 690;
 Best Local Similarity 80.7%; Pred. No. 6.4e-210;
 Matches 556; Conservative 54; Mismatches 75; Indels 4; Gaps 2;
 QY 1 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTDEVSKREGINAEQIVIKITDQGYVT 60
 Db 2 AYELGLHQATVKNRRVSYIDGKQATOKTENLTDEVSKREGINAEQIVIKITDQGYVT 61
 QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIDVNEVKGGVYIKVDGKVVYVLKDA 120
 Db 62 SHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIDVNEIKGVYIKVNGKYYVLKDA 121
 QY 121 AHADNVRTKEINRQKQHSQHREGGTPRNGGAVALARSQGRYTTDDGYIFNASDIIEDT 180
 Db 122 AHADNVRTKEINRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASDIIEDT 181
 QY 181 GDAYIVPHGDHYHYIPKKNELGASLAAAEAFLSGRGNLSNRTYRQNSDNTSRNWPVS 240
 Db 182 GDAYIVPHGDHYHYIPKKNELGASLAAAEAFLSGRGNLSNRTYRQNSDNTSRNWPVS 241
 QY 241 VSNPQTNTNTSNNSTNSQASQNSNDISLLKQLYKPLSORHVESDGLVDFPAQITSR 300
 Db 242 VSNPQTNTNTSNNSTNSQASQNSNDISLLKQLYKPLSORHVESDGLVDFPAQITSR 301
 QY 301 ARGVAVPHGDHYHYIPYSQMSLEBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
 Db 302 ARGVAVPHGNHYHFIPEQMSLEKRIARIIPLYRSNHWVPDSRPEPSPQPTPEPSPS 361
 QY 361 POPAPNLK-IDSN--SSLSYQLVRKVGEGYVFEKGISRYVFAKDLPSKVNLKSLK 416
 Db 362 POPAPNQPPAPNPIDKLVKAEVRKVGVDGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
 QY 417 KQESVSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLND 476
 Db 422 KQESLSHKLGAKTDLPSDDREFYKAYDILLARHQDLNDKGRQVDFEALDNLRLERKD 481
 QY 477 ESTNKEKLVDDLLAFAPITHTPERLGKPSQIEYTEDVIRIAQLADKYTTSDGYIFDEHD 536
 Db 482 VSDKVKLVDDLLAFAPITHTPERLGKPSQIEYTEDVIRIAQLADKYTTSDGYIFDEHD 541
 QY 537 IISDEGDAYVTPHMGSHWIGKSDLSDEKVAQAAYTKEKGLPSPDADVKANPTGDSA 596
 Db 542 ITSDEGDAYVTPHMGSHWIKKDSLSAEABRAAAQAYAKEKGLTPSTDHDSGNTAKGA 601
 QY 597 RAIYNRVKGEKPIPLVRLPYMVEHTVEKNGNLIIPHKDHYHNKFAWDDHTYKAPNGY 656
 Db 602 EAIYNRVKAAKVPDLRMPYNLQYTVVEKNGSLIIPHYDHYHNKFEWDEGLYEAPKY 661
 QY 657 TLEDLFATIKYVVEHDPDERPHSDNGWGNA 685
 Db 662 TLEDLLATVKYVVEHPNRPHERPHSDNGFGNA 690

RESULT 16
 AAY91939
 ID AAY91939 standard; protein; 826 AA.
 XX
 AC AAY91939;
 XX
 DT 19-JUL-2000 (first entry)
 DE S. pneumoniae 92 kDa human C3-degrading protein.
 KW Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
 KW inhibitor; inflammation; organ rejection; xenotransplantation.
 OS Streptococcus pneumoniae.
 XX
 PN WO200017370-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 24-SEP-1999; 99WO-US022362.

XX	24-SEP-1998;	98US-0101736P.		Qy	536	DIISDEGDVYTPHMGHSHWIGKDSKDKVAAQYTKKGLPPSPDADVKANPTGDS	595
PR	31-MAR-1999;	99US-00283094.		Db	547	DITSDEGDVYTPHMTSHWIKDLSGEAERAAQAAYAKKGLTPPSTDHODSGNTRAKG	606
XX	(MINU) UNIV MINNESOTA.			Qy	596	AAAIYRVKGEKRIPLARLPMVVEHTVEVKNGLIIPHKDHYHNIKPAWFDDHTYKAPNG	655
PA	(AMCY) AMERICAN CYANAMID CO.			Db	607	ABAIYRVKAAKVPLDORMPNLQYTVVEVKNGLIIPHYDHYHNIKPFWFDEGLYEAPKG	666
XX	Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;			Qy	656	YTLDELATIKYVVEHPDERPHSDNGMGNASEHVLGKHSDSDPNKPFKADEE-----P	709
XX	WPI; 2000-283594/24.			Db	667	YTLDELATIKYVVEHPDERPHSDNGMGNASEHVLGKHSDSDPNKPFKADEE-----P	726
XX	N-PSDB; AAA08557.			Qy	710	VEET-----PABEVEVQVETEKVEAQIKAEVLLAKVTDSSL	746
XX	Isolated polypeptide is used to stimulate immune system and immunize or			Db	727	EETPREKPSQKPEPKPTTEEPSEPESEEPQVETEKVEEKLUREAEDLLGKIQDPII	786
XX	treat a mammalian subject against Streptococcus pneumoniae infection or			Qy	747	KANATETLAGLNLLTQIMDNNSIMAEKLLALLKGS	785
XX	colonization.			Db	787	KSNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKES	825
XX	Claim 8; Page 55-57; 63pp; English.						
XX	The present sequence, isolated from Streptococcus pneumoniae, is a human						
CC	C3-degrading protein of about 92 kDa. This sequence may encompass a						
CC	smaller, approximately 20 kDa protein (see AA91938), also having human						
CC	C3-degrading activity. The DNA sequences (AAA08556-57) can be used for						
CC	producing an immune response to Streptococcus pneumoniae in a mammal.						
CC	Antibodies against the proteins can be used to inhibit S. pneumoniae-						
CC	mediated C3 degradation. C3-mediated inflammation and rejection in						
CC	xenotransplantation can be inhibited by expressing the nucleic acid						
CC	sequences on the surface of an organ of an animal. In particular, the						
CC	polypeptides are useful for stimulating the immune system and are						
CC	effective to immunize or treat a mammalian subject against Streptococcus						
CC	pneumoniae infection or colonization						
XX	Sequence 826 AA;						
SQ	Query Match	67.5%; Score 2812; DB 3; Length 826;					
	Best Local Similarity	67.4%; Pred. No. 9e-199;					
	Matches	552; Conservative 81; Mismatches 138; Indels 48; Gaps 8;					
Qy	1 SYELGLYQARTV-KENNRVSYIDGKAQTKENTLTPDEVSKREGINAEQIVIKITDQGVV	59					
Db	21 SYELGRHQAGVQKESNRVSYIDGQAGKAENLTPDEVSKREGINAEQIVIKITDQGVV	80					
Qy	60 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIYNEIKGGVYKVDGKYVYVKD	119					
Db	81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIYNEIKGGVYKVDGKYVYVKD	140					
Qy	120 AAHADNVRTKEEINRQKQHSQREGGTPTNDGAVALARSGHYTTDDGYIFNASDIIE	179					
Db	141 AAHADNVRTKEEIKRQKQHSQREGGTPTNDGAVALARSGHYTTDDGYIFNASDIIE	198					
Qy	180 TGDAYIVPHGDHYHYIPKNEISASELAFAELSGRNLNSRTYRRQNSDNTSRTNWVP	239					
Db	199 TGDAYIVPHGDHYHYIPKNEISASELAFAELSGRNLNSRTYRRQNSDNTSRTNWVP	247					
Qy	240 SVNPGTNTNTSNNNTNSQASQNDIDSLKQLYKLPLSRHVESDGLVFPQAITSR	299					
Db	248 NPAQPLSENHNTVPTTHQ-NQGENISLLRELAKPLSERHVESDGLVFPQAITSR	306					
Qy	300 TARGVAVPHGDHYHYHYFIPYQMSLEERARIIPLRYSNHWVDSRPEQSPQPTPEPSP	359					
Db	307 TARGVAVPHGDHYHYHYFIPYQMSLEERARIIPLRYSNHWVDSRPEQSPQPTPEPSP	366					
Qy	360 GPAPNLIK-IDSN---SSIVSOLVRKVGEGYFEEKIGSRVFAKPLSETVKNLESKL	415					
Db	367 SPOAPNPQAPSNPIDEKLVKAVRKGVGEGYFEEKIGSRVFAKPLSETVKNLESKL	426					
Qy	416 SKQSVSHLTITAKENAVRQDFYDKAYNLLTEAHKALFKXNKGNSDFQALDKLRLN	475					
Db	427 AKQESLSHKLGAKKTDLPSSDRFPYKAYDILLARIHQDLNDKGRQVDFEALDNLRLK	486					
Qy	476 DESTNKEKLVLDLIAFLAPIHTHPRKLPNSQIETDEVRIAQLADKYTSDGYIDEH	535					
Db	487 DVPSDKVKLVLDLIAFLAPIHTHPRKLPNSQIETDEVRIAQLADKYTSDGYIDEH	546					

RESULT 17
ASU01597
ID ABU01597 standard; protein; 819 AA.
AC ABU01597;
XX
XX 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain protein from coding region #1173.
XX
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB002163.
XX
XX 27-MAR-2001; 2001GB-00007658.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX N-PSDB; ABX06885.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
XX Claim 1; SEQ ID NO 2346; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AS956454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence

CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ

Sequence 819 AA;

Query Match 67.2%; Score 2797.5; DB 6; Length 819;
Best Local Similarity 67.0%; Pred. No. 1.1e-197;
Matches 546; Conservative 84; Mismatches 136; Indels 47; Gaps 7;

QY 1 SYELGLYQA-RVKNRVSVDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGV 59
DB 21 SYELGRYQAGQDKKESNRVAVIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGV 80
QY 60 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIVNEVKGYVVKVPGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIVNEVKGYVVKVPGKYVYLKD 140
QY 120 AAHADNVRTKEEIKRQKQESHREGGTPRNDGVALARSQRYTTDDGYIFNADIIED 179
DB 141 AAHADNVRTKEEIKRQKQESHREGGTPRNDGVALARSQRYTTDDGYIFNADIIED 197
QY 180 TGDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTVRQNSDNTSRTWVP 239
DB 198 TGDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTVRQNSDNTSRTWVP 246
QY 240 SVSNPGTNTNTSNNTSNQASQNSDIDSLKOLYKPLSORHVESDGLVFPDPAQITSR 299
DB 247 NPQAPRLSENHLTVTTTYHQ-NQGENISSLLRELYAKPLSERHVESDGLVFPDPAQITSR 305
QY 300 TARGVAVPHGDHYHYFIPYSQMSLEERARIIPLYRVSNEHWVDSRPEQSPQPTPEPSP 359
DB 306 TARGVAVPHGNHVFIPYQMSLEEKRIARIIPLYRVSNEHWVDSRPEQSPQPTPEPSP 365
QY 360 GPQAPNKLIDSSSLVSLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSQKE 419
DB 366 SPQAPASNPIDE-EKLVKAERVKVGDCGVFEENGVSRYIPAKOLSASTAGIADSKLAKQE 423
QY 420 SVSHGTLTAKENAVPRDQEFYKAYNLLTAHKAFLFNKGRNSDFQALDKLERLNDEST 479
DB 424 SLHGKLTGKTIDLPSSDREYFNKAYDILARIHQDLNDKGRQVDFEALDNLRLKDVSS 483
QY 480 NKEKLVDDLLAFAPLTHPERLKPNSOIEYTVDEVRILQADKYTTSDGYIFDEHDIIS 539
DB 484 DKVKLVEDILAFAPLTHPERLKPNSOIEYTVDEVRILQADKYTTSDGYIFDEHDIIS 543
QY 540 DEGDAYVTPHMGHSHWIGKDSLSKQKVAQAAYTKBKGLPSPDADVKANPTGDSAAAI 599
DB 544 DEGDAYVTPHMTSHWIKDLSSEARAAQAAYAKKGLTPPTDQDSNGTAKAEAI 603
QY 600 YNRVKEKRIPLVRLPYMVEHTEVKNGLIIPHKDHYHNIKFAWPDHHTYKAPNGYTLE 659
DB 604 YNRVKAASKVPLDRMPYNLQYTVVEVKNGLIIPHYDHYHNIKEFWDEGLYKAPNGYTLE 663
QY 660 DLFATIKYVVEHDPDRPHNSDNGMNASEHVLGKDHSEDPNKNFKADEE-----PVET 713
DB 664 DLLATVYVVEHDPDRPHNSDNGMNASEHVLGKDHSEDPNKNFKADEE-----PVET 723

QY 714 -----PAPPEVPQVETEKVEAQLKEAEVLLAKVTDSSLKANA 750
DB 724 PREEKQSEKPSKPTPEEPSESPERQVETEKVEKLEAEADLLKIQDPIKNSA 783
QY 751 TETLAGRLNLLTQIMDNNSINAEAKLLALLKGS 785
DB 784 KETLGLKNLLFGTQDNNITMAEAKLLALLKES 818
RESULT 18
AAAY81662
ID AAAY81662 standard; protein; 827 AA.
XX
AC AAAY81662;
XX
DT 24-MAY-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID311.
XX
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006737-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB002451.
XX
PR 27-JUL-1998; 98GB-00016337.
PR 19-MAR-1999; 99US-0125164P.
XX
PA (MICR-) MICROBIAL TECHNIQS LTD.
XX
PI Gilbert CFG, Hansbro PM;
XX
DR MPI; 2000-195300/17.
XX
PT New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing or
PT inhibiting expression of the protein.
XX
PS Claim 2; Page 99-100; 108pp; English.
XX
CC AAAY81501 to AAAY81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AAAY81501 to AAAY81679 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties. The
CC protein sequences, and fragments of them, are useful as immunogens and/or
CC antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AAAY81501 to AAAY81679 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 827 AA;

Query Match 67.1%; Score 2795; DB 3; Length 827;
Best Local Similarity 67.0%; Pred. No. 1.1e-197;
Matches 549; Conservative 83; Mismatches 139; Indels 48; Gaps 8;

QY 1 SYELGLYQA-RVKNRVSVDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGV 59
DB 21 SYELGRHQAQDKKESNRVAVIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGV 80
QY 60 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIVNEVKGYVVKVPGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISSELLMKDPNKLKDEIVNEVKGYVVKVPGKYVYLKD 140

Db 544 DEGDAYVTPHMTSHWIKKDSLSAEARAAQAAYAEKGLTPPSTDHODSGNTEAKGAEAI 603
 QY 600 YNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNKIFKFAWDDHTYKAPNGYTYLE 659
 Db 604 YNRVKAARKVPLDRMPYNLQYTVKNGSLIIPHVDYHNKIFKFEWDEGLYEAAPKGYTYLE 663
 QY 660 DLFATIKYVVEHPDERPHSNDGWNASRHLVKGKDHSDPNKFKADEE-----PVEET 713
 Db 664 DLLATVKYVVEHPNERPHSDNGFGNASHVQVQNKNGQADTNQTEKPSSEKQTEKPEET 723
 QY 714 -----PAPEVPOVETEKVEAQLKEAEVLLAKVTDSS 745
 Db 724 PREEKPOSEKPSKPTEPEESPESEPEQVETEKVEKLEAREADLLGKIQNP 764
 QY 751 TETLAGLNNTLQIMDNNSIMAEKLLALLKGS 785
 Db 784 KETLTGLKNLLFGTQDNNTIMAEKLLALLKES 804

RESULT 20
 AAB12764
 ID AAB12764 standard; protein; 805 AA.
 AC AAB12764;
 XX
 DT 21-NOV-2000 (first entry)
 XX Streptococcus pneumoniae strain JNR7/87 BVH-11 protein antigen.
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX Streptococcus pneumoniae.
 OS
 PN WO200039299-A2.
 XX 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 XX
 PS Disclosure; Fig 12; 106pp; English.
 XX

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein antigen, from the present invention

Sequence 805 AA;
 Query Match 66.9%; Score 2786; DB 3; Length 805;
 Best Local Similarity 66.3%; Pred. No. 7.3e-197;
 Matches 544; Conservative 88; Mismatches 136; Indels 52; Gaps 7;
 1 SYELGLYQA-RTVKENNRVSYIDGKQATKENTLTPDEVSKREGINAEQIVIKITDQGV 59
 2 SYELGRHQAGQDKKESNRVAYIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGV 61

QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGYVIVKDGKYYVYLKD 119
 Db 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKSDIVNEIKGGYVIVKNGKYYVYLKD 121
 QY 120 AAHADNVRTKEEINROKOEHSQHREGGTPRNDGVALARSQGRYTTDDGYIFNASDIIED 179
 Db 122 AAHADNVRTKEEIKRQKQERSHNHNS---RADNVAARAQGRYTTDDGYIFNASDIIED 178
 QY 180 TGDYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTVRRQNSDNTSRNTVWP 239
 Db 179 TGDYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTVRRQNSDNTSRNTVWP 227
 QY 240 SVSNPGTNTNTNSNTNSQASQSDNDISLLKOLYKLPQSRHVESDGLVDFDPAQITSR 299
 Db 228 NPAQFRLSENHNLTVTPTYHQ-NQGENISLLRELYAKPLSERHVESDGLVDFDPAQITSR 286
 QY 300 TARGVAVPHGDHYHYFIPYSOMSELEERTARIPIURYSNHWPDSRPEQSPQTPPEPSP 359
 Db 287 TARGVAVPHGNHYHFIPYEQNSELEKRIARIPIURYSNHWPDSRPEPSPQTPPEPSP 346
 QY 360 GPQAPNLKIDSNSLSVQLVRKVGEGYVPEEKGISRYVFAKOLPSETVKNLEKSKOE 419
 Db 347 SPQAPSNPID-EKLVKEAVRKVGDSYVFEENGVSRYIPAKDLSAETAAGIDSKLAKOE 404
 QY 420 SVSHTLTAKENVAPRQOEYOKAYNLLTEAKALFXNKGNSDFQALDKLLERLNDEST 479
 Db 405 SLSHKLGAKKTDLPSSDRFYNKAYDILLARIHQDLNDKGRQVDFEALDNLRLKDVSS 464
 QY 480 NKEKLVDDLLAFAPITHPERLGPNSOIBYTEDEVRIAGIADKYTTSDGYIFDEHDIIS 539
 Db 465 DKVKLVDDIILAFAPINRPERLGPNAQIITYTDBEIQVAKLAGKYTTEDGYIFDPRDITS 524
 QY 540 DEGDYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEGILPPSPDADVKANPTGDSAAAI 599
 Db 525 DEGDYVTPHMTSHWIKKDSLSAEARAAQAAYAEKGLTPPSTDHODSGNTEAKGAEAI 584
 QY 600 YNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNKIFKFAWDDHTYKAPNGYTYLE 659
 Db 585 YNRVKAARKVPLDRMPYNLQYTVKNGSLIIPHVDYHNKIFKFEWDEGLYEAAPKGYTYLE 644
 QY 660 DLFATIKYVVEHPDERPHSNDGWNASRHLVKGKDHSDPNKFKADEE-----PVEET 713
 Db 645 DLLATVKYVVEHPNERPHSDNGFGNASHVQVQNKNGQADTNQTEKPEEKPEET 704
 QY 714 -----PAPEVPOVETEKVEAQLKEAEVLLAKVTDSS 745
 Db 705 PREEKPOSEKPSKPTEPEESPESEPEQVETEKVEKLEAREADLLGKIQNP 764
 QY 746 LKANATETLAGLNNTLQIMDNNSIMAEKLLALLKGS 785
 Db 765 IKSNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKES 804

RESULT 21
 AAB12755
 ID AAB12755 standard; protein; 820 AA.
 XX
 AC AAB12755;
 XX
 DT 21-NOV-2000 (first entry)
 XX Streptococcus pneumoniae strain JNR7/87 BVH-11-2 protein antigen.
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.

[illegible]

Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY	1	SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTPEVSKREGINAEOIVIKITDQGV	59
Db	21	SYELGRHQAGGVKKESNRVSYIDGQAGKAEINTPEVSKREGINAEOIVIKITDQGV	80
QY	60	TSHGDHYHYNGKVPYDAIISEELLMKDPNYKLKDEDIVNEVKGYVIVKDGKYYVYLKD	119
Db	81	TSHGDHYHYNGKVPYDAIISEELLMKDPNYQLKSDIVNEIKGGYVIVKDGKYYVYLKD	140
QY	120	AAHADNVRTKEEINRQKEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIITD	179
Db	141	AAHADNIRTKKEIKRQKEHSHNHGGGS--NDQAVVAARAGRYTTDDGYIFNASDIITD	198
QY	180	TGDYIVPHGDHYHYVIPKNEISASLEAAAEAFLSGRGNLSRSTVRRQNSDNTSTNWVP	239
Db	199	TGDYIVPHGDHYHYVIPKNEISASLEAAAEAYWG-----KQGRSPSSSSSYNA	247
QY	240	SVSNPGTTNTNTSNNSNTNSQASQNDISLLKQLYKLPISORHVESDGLVPDPAQITSR	299
Db	248	NPAQPRLSNHNLTVTPTVHQ--NOCENISLLRELYAKPLSERHVESDGLVPDPAQITSR	306
QY	300	TARGVAVPHGDHYHYPIPYQMSSELRARIIPLYRSNHWVPDSRPQSPQPTPEPSP	359
Db	307	TARGVAVPHGNHYHPITYEQMSLEKRIARIIPLYRSNHWVPDSRPQSPQSPQPTPEPSP	366
QY	360	GPOAPNPK--IDSN--SSLSVQLRVKGVGVFEKIGISRVVFAKDIIPSEVFNKLESKL	415
Db	367	SPQAPNPQPPSPNPIDEKLVEARVKVGDGVFEENGVSRIIPAKDISAETAAIGDSKL	426
QY	416	SKQESVSHLTAKKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDQFQALDKLLERLN	475
Db	427	AKQESLSHKLGAKKTDLPSSDREFYNKAYDILLARIHQDILLDNKGRQVDFEALDNLLERLK	486
QY	476	DESTNKEKLVDDLAFAPITPERLKGNSQIETDEVRIAQLADKVTTSIDGYIFDSEH	535
Db	487	DVPSDKVLVDILAFAPIRPERLKGPNQAQITTDDEIQVAKLAGKVTTEDEGYIFDPR	546
QY	536	DIISDEGDYAVTPHMGHSHWIGKDSLDEKVAQAAYTKEKGLTPSPDADVKANPTGDS	595
Db	547	DITSDEGDYAVTPHMTSHWIKDLSSEARAAQAYAKEKGLTPTSDHQDSNGTEAGK	606
QY	596	AAAIYNRVKGEKRIPIVLRLPYMVVEHTVEVKNGLIIPHDKDHNHINKAFWPDHHTYKAPNG	655
Db	607	AEAIYNRVKAACKVPLDRMPYNLYQTVVEVKNGLIIPHYDHYHNIKFEWFDEGLYAPKG	666
QY	656	YTIEDLPALKYVVEHPDERPHSDNGWGNASHEVLGK-----KDHS-----	697
Db	667	YTIEDLLATKYVVEHPDERPHSDNGWGNASDHVRKNKVDQDSKPDDEKHEHDEVEPTHP	726
QY	698	-----DPNKNFKADPEPVBETPAEPEVPOVETEKVEAQLKAEAVLL	738
Db	727	ESDEKENHAGLNPSADNLYKPSDTEETEEAEADTTDEAEIIPQVENSVINAKIADAELL	786
QY	739	AKVTDSSLKANATETLAGLNRLNLTQIOMNNSTMAEAKLLALLKGSNPSSV	790
Db	787	EKVTDPSIRQAMETTLGLKSSLLGTKNNTTISAEVDSLALLLKESQAPAI	838
RESULT 24			
ID	AAB12765		
XX	AAB12765 standard; protein; 807 AA.		
AC	AAB12765;		
XX	21-NOV-2000 (first entry)		
DT	Streptococcus pneumoniae strain SP63 BVH-11 protein antigen.		
DE	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;		
XX	prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;		
KW	otitis media; pneumonia; immunisation; bactericidal.		
KW	Streptococcus pneumoniae.		
OS			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 06:41:29 ; Search time 28 Seconds
(without alignments)
2734.588 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTVKENRVSY.....KLALLKGSNPFSSVSKKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

PIR_78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4163	100.0	802	E95136	conserved domain p
2	4159	99.9	828	E98004	hypothetical prote
3	3224	77.4	855	D98004	histidine Motif-Co
4	2797.5	67.2	819	E95136	conserved domain p
5	2772	66.6	839	G95115	conserved hypothet
6	2733	65.6	853	C97985	conserved hypothet
7	1246	29.9	1039	H95115	conserved hypothet
8	1243	29.8	1039	D97985	hypothetical prote
9	934	22.4	822	T46758	hypothetical prote
10	243	5.8	182	E97985	hypothetical prote
11	192.5	4.6	1390	T44004	hypothetical prote
12	176.5	4.2	1271	A45555	hypothetical prote
13	172	4.1	2481	D90011	trfA protein - sli
14	171	4.1	1233	S56271	Glutamate rich pro
15	168.5	4.0	1043	D84900	FntB protein [limp
16	165	4.0	891	C97985	hypothetical prote
17	161	3.9	1420	A44361	hypothetical prote
18	159	3.8	2810	T22298	hypothetical prote
19	158	3.8	1650	T18444	hypothetical prote
20	157	3.8	910	S73361	dnaJ homolog prote
21	157	3.8	1658	T22298	hypothetical prote
22	157	3.8	2195	S61103	hypothetical prote
23	156.5	3.8	1495	T48429	SEC16 protein - ye
24	156	3.7	519	C84598	hypothetical prote
25	154.5	3.7	1338	T18416	probable bZIP tran
26	154.5	3.7	1463	T30290	hypothetical prote
27	154.5	3.7	1516	E71619	AA5 surface protei
28	154.5	3.7	3924	S37431	RAD2 endonuclease
29	154	3.7	749	S23467	ankyrin 2, neurona
					probable long-chai

RESULT 1

C95136

conserved domain protein SP1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: C95136
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide

ALIGNMENTS

30	151.5	3.6	1163	2	A36685	205K microtubule-a
31	150.5	3.6	1176	2	S66771	hypothetical prote
32	150	3.6	1132	2	B82538	ribonuclease E XF2
33	150	3.6	2401	2	T38676	phoptry protein -
34	149.5	3.6	1175	2	S51005	protein-tyrosine-p
35	149	3.6	1274	2	A89959	hypothetical prote
36	149	3.6	2829	2	A42771	reticulocyte-bindl
37	149	3.6	4688	2	F82885	hypothetical prote
38	149	3.6	6713	2	B89921	hypothetical prote
39	148.5	3.6	1785	2	T21558	hypothetical prote
40	148.5	3.6	1959	2	AG1085	hypothetical prote
41	148.5	3.6	1969	2	T38495	hypothetical prote
42	148	3.6	980	2	E71606	hypothetical prote
43	148	3.6	1230	2	T22458	hypothetical prote
44	148	3.6	1856	2	C95008	immunoglobulin A1
45	147	3.5	700	2	A56976	transfer complex p
46	147	3.5	1269	2	F84730	probable myosin he
47	147	3.5	1929	2	T15559	hypothetical prote
48	146.5	3.5	2004	2	F51133	immunoglobulin A1
49	146	3.5	821	2	S67087	hypothetical prote
50	145.5	3.5	1073	2	S14032	hypothetical prote
51	145.5	3.5	1085	2	T38378	kinesin-related pr
52	145.5	3.5	1240	2	S52734	kinesin-like prote
53	144.5	3.5	934	2	T47546	hypothetical prote
54	144.5	3.5	4152	2	T31102	protein kinase-lik
55	144.5	3.5	4919	2	T31105	filamentous hemagg
56	143.5	3.4	900	2	C64232	hypothetical prote
57	142.5	3.4	875	2	S70115	alanine-tRNA ligas
58	142.5	3.4	1066	2	T45283	ZIP1 protein - yea
59	142.5	3.4	1066	2	T41099	growth polarity ma
60	142.5	3.4	1979	2	C71622	staurosporine targ
61	142.5	3.4	5170	2	T15348	hypothetical prote
62	142	3.4	1257	2	T00486	hypothetical prote
63	142	3.4	1744	2	JH0720	serine/threonine-s
64	142	3.4	1822	2	S33441	canabin - African
65	141.5	3.4	490	2	S52830	EF protein - Strep
66	141.5	3.4	1315	2	T28679	HMS1 protein - yea
67	141.5	3.4	1928	2	S46773	fibrinogen-binding
68	141.5	3.4	2314	1	A46151	myosin heavy chain
69	141.5	3.4	2748	2	S57976	protein-tyrosine-p
70	141.5	3.4	3498	2	T22330	nuclear migration
71	141	3.4	914	2	B48086	hypothetical prote
72	141	3.4	1190	2	S47536	translation initia
73	141	3.4	1280	2	T42514	oxysterol-binding
74	141	3.4	1332	2	S41552	kinase anchor prot
75	141	3.4	2288	2	T29999	probable transcrip
76	141	3.4	3890	2	C89921	hypothetical prote
77	140.5	3.4	507	2	S05542	hypothetical prote
78	140.5	3.4	719	2	A81358	hypothetical prote
79	140.5	3.4	2231	2	D71870	hypothetical prote
80	139.5	3.3	1535	2	T49042	hypothetical prote
81	139.5	3.3	1609	1	MMHUB2	hypothetical prote
82	139.5	3.3	4549	2	T20771	laminin gamma-1 ch
83	139.5	3.3	4667	2	T20774	hypothetical prote
84	139	3.3	1247	2	C89583	protein K07E3.1 [1
85	139	3.3	1385	2	D89824	hypothetical prote
86	139	3.3	1957	2	T38077	hypothetical prote
87	138.5	3.3	1703	2	S15047	hypothetical coile
88	138.5	3.3	2269	2	T28677	SNF2 protein - yea
89	138	3.3	1639	2	S05603	thoptry protein -
90	138	3.3	2295	2	B71621	major merozoite su
						probable membrane

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: C95136
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-802 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:gl4972655; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 C:Gene: *phA*
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 100.0%; Score 4163; DB 2; Length 802;
 Best Local Similarity 99.9%; Pred. No. 1.7e-219;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRVSIDGKQATQKTENLTDEVSREGINAQIVIKITDQGYT 60
 Db 7 SYELGLYQARTVKNRVSIDGKQATQKTENLTDEVSREGINAQIVIKITDQGYT 66

QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNKLKDEDIVNEVGKGYVVKDGYVYVLKDA 120
 Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNKLKDEDIVNEVGKGYVVKDGYVYVLKDA 126

QY 121 AHADNVRTKEENRQKHSQREGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180
 Db 127 AHADNVRTKEENRQKHSQREGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 186

QY 181 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSDNTSRTNWVPS 240
 Db 187 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSDNTSRTNWVPS 246

QY 241 VSNPGTTNTNNSNTNSQASNDISLLKQLYKPLSQRHVESDGLVDFPAQITTSRT 300
 Db 247 VSNPGTTNTNNSNTNSQASNDISLLKQLYKPLSQRHVESDGLVDFPAQITTSRT 306

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
 Db 307 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 366

QY 361 PQAPNKLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKOLPSETVKNLESKLSQES 420
 Db 367 PQAPNKLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKOLPSETVKNLESKLSQES 426

QY 421 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 Db 427 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 486

QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 540
 Db 487 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 546

QY 541 EGDAYVTPHMGSHHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADVKANPTGDSAAIY 600
 Db 547 EGDAYVTPHMGSHHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADVKANPTGDSAAIY 606

QY 601 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNIFKAFWDDHTTYKAPNGYTTLED 660
 Db 607 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNIFKAFWDDHTTYKAPNGYTTLED 666

QY 661 LFATIKYVVEHPDRPHNSDGNWSEHVLGKDHSEDPNKNKFADEPVEETPAEPEVP 720
 Db 667 LFATIKYVVEHPDRPHNSDGNWSEHVLGKDHSEDPNKNKFADEPVEETPAEPEVP 726

QY 721 QVTEKVEAQLKAEVLLAKVTDSLSKANATETLAGLNLLTQIMDNNSIMAEAEKLLA 780
 Db 727 QVTEKVEAQLKAEVLLAKVTDSLSKANATETLAGLNLLTQIMDNNSIMAEAEKLLA 786

QY 781 LLKGSNFPSSVSKEKIN 796

Db 787 LLKGSNFPSSVSKEKIN 802

RESULT 2
 E98004
 hypothetical protein *phA* [imported] - *Streptococcus pneumoniae* (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: E98004
 R:Hoekins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E.
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M.
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: E98004
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-828 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:gl5458683; GSPDB:GN00174
 C:Genetics:
 C:Gene: *phA*
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 99.9%; Score 4159; DB 2; Length 828;
 Best Local Similarity 99.7%; Pred. No. 2.9e-219;
 Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRVSIDGKQATQKTENLTDEVSREGINAQIVIKITDQGYT 60
 Db 33 SYELGLYQARTVKNRVSIDGKQATQKTENLTDEVSREGINAQIVIKITDQGYT 92

QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNKLKDEDIVNEVGKGYVVKDGYVYVLKDA 120
 Db 93 SHGDHYHYNGKVPYDAIISEELMKDPNKLKDEDIVNEVGKGYVVKDGYVYVLKDA 152

QY 121 AHADNVRTKEENRQKHSQREGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180
 Db 153 AHADNVRTKEENRQKHSQREGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 212

QY 181 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSDNTSRTNWVPS 240
 Db 213 GDAYIVPHGDHYHYIPKNELASELAAAFISGRGNLSNRYRQNSDNTSRTNWVPS 272

QY 241 VSNPGTTNTNNSNTNSQASNDISLLKQLYKPLSQRHVESDGLVDFPAQITTSRT 300
 Db 273 VSNPGTTNTNNSNTNSQASNDISLLKQLYKPLSQRHVESDGLVDFPAQITTSRT 332

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
 Db 333 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 392

QY 361 PQAPNKLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKOLPSETVKNLESKLSQES 420
 Db 393 PQAPNKLKIDSNSLSVQLVRKVGEGVPEEKISRYVFAKOLPSETVKNLESKLSQES 452

QY 421 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
 Db 453 VSHLTAKKENVAPRQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 512

QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 540
 Db 513 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIQAQADKYTSDGYIFDEHDIISD 572

QY 541 EGDAYVTPHMGSHHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADVKANPTGDSAAIY 600
 Db 573 EGDAYVTPHMGSHHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADVKANPTGDSAAIY 632

QY 601 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNIFKAFWDDHTTYKAPNGYTTLED 660
 Db 633 NRKVGKRIPLVRLPYMVEHTVEKGNLIIIPKDHVHNIFKAFWDDHTTYKAPNGYTTLED 692

QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKADREPPVEETPAEPVP 720
 Db 693 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKHSDPNKNFKADEPVEETPAEPVP 752
 QY 721 QVETEKVBAQIKEARVLLAKVTDSSLSKANATETLAGLRNNLTQIMDNNSIMAAEKLKLA 780
 Db 753 QVETEKVBAQIKEARVLLAKVTDSSLSKANATETLAGLRNNLTQIMDNNSIMAAEKLKLA 812
 QY 781 LLKGSNPSSVSKEKIN 796
 Db 813 LLKGSNPSSVSKEKIN 828

RESULT 3
 D98004
 histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
 C;Species: Streptococcus pneumoniae
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C;Accession: D98004
 R;Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 e, R.; Lelanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, F.; McAnren, S.; M
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:11544234
 A;Accession: D98004
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-855 <KUR>
 A;Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:g15458682; GSPDB:GN00174
 C;Genetics:
 A;Gene: phpA
 C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 77.4%; Score 3224; DB 2; Length 855;
 Best Local Similarity 75.0%; Pred. No. 3.2e-168;
 Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;

QY 1 SYELGLYQARTKVENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
 Db 32 AYELGLHQATQKVENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 91
 QY 61 SHGDHYHYNGKVPYDAIIESELLMKDPNYKLKDEDIVNEVKGYYVVKDGYVYVYKDA 120
 Db 92 SHGDHYHYNGKVPYDAIIESELLMKDPNYKLKDEDIIEIKGGYVVKDGYVYVYKDA 151
 QY 121 AHADNVRTKEEINRQKQESHQREGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
 Db 152 AHADNVRTKEEINRQKQESHQREGTTPNDGAVALARSGRYTTDDGYIFNASDIIEDT 211
 QY 181 GDAYIVPHGDHYHYIPKNELSAELAAEAFUSGRGNLSNSTRYRRONSNDTSTRNWWVPS 240
 Db 212 GDAYIVPHGDHYHYIPKNELSAELAAEAFUSGRGNLSNSTRYRRONSNDTSTRNWWVPS 271
 QY 241 VSNPQTNTNTSNNTNSQASQSDNDISLLKQYKLPQSRHVSDDGLVDPQAITSRT 300
 Db 272 VSNPQTNTNTSNNTNSQASQSDNDISLLKQYKLPQSRHVSDDGLVDPQAITSRT 331
 QY 301 ARGVAVPHGDHYHYIPYQSMGELEERARIIPLYRSHNWPDSPRPEQSPQTPPEPSPG 360
 Db 332 ANGVAVPHGDHYHYIPYQSLPLEEKARIIIPLYRSHNWPDSPRPEQSPQTPPEPSPG 391
 QY 361 PQAPNPK- IDSN---SSIVSLQVKKVGGYVVEEKGISRYVFAKDLPSRTVKNLESKLS 416
 Db 392 PQAPNPOAPSNPIDKLVKEARVKGVDGYVVEENGVPVIPAOKDLSAETAAGIDSKLA 451
 QY 417 KQESVSHLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLND 476
 Db 452 KQESLSHLGAKKTDLPSSDREFYNKADLLARITHQDLDNKGQVDFEALDNLLERLND 511
 QY 477 ESTNKEKLVDDLLAFAPITPHERLIGKPNQSIETYETDEVRILAQLADKYTTSDGYIFDHD 536

Db 512 VSDKVKLVDDILAFAPIRPERLIGKNAQITVTDDEIQVAKLAGKYTTEDGYIFDPRD 571
 QY 537 IISDEGDAYVTPHMGSHWICKDSIDKEKVAQAAYTEKGLPSPADVKANPTGDSA 596
 Db 572 ITSDEGDAYVTPHMGSHWICKDSIEABRAAAQAYAKEKGLTPPSTDHODSGMTEAKGA 631
 QY 597 AAIYNRVKGEKRIPLVRLPYMVEHTVEVXGNLLIIPHKDHYHNKIFAWDFDHTYKAPNGY 656
 Db 632 EAIYNRVKAACKVPLDRMPYNLQTVVEVXGSLIIPHYDHYHNKIFEFDFGLFEAPKGY 691
 QY 657 TLEDLFATIKYVVEHPDERPHSDNGWGNASHVVGKDHSDPNKNFKADEE-----PV 710
 Db 692 SLEDLLATVYVVEHPDERPHSDNGWGNASHVVGKDHSDPNKNFKADEE-----PV 751
 QY 711 EET-----PAPPEVPOVETEKVEAQLKEAEVLLAKVT 742
 Db 752 EETPREEKPOSEKPEPKTEEPSESPSESEPOVETEKVEKLEKREADELLGKIQ 811
 QY 743 DSSLKANATETLAGLRNNLTQIMDNNSIMAAEKLKLAALLKGS 785
 Db 812 NPIIKSNAKETLTGKXNLLFGTQDNNTIMAAEKLKLAALLKES 854

RESULT 4
 B95136
 conserved domain protein SP1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
 C;Species: Streptococcus pneumoniae
 C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C;Accession: B95136
 R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A;Reference number: A95000; MUID:21357209; PMID:11463916
 A;Accession: B95136
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-819 <KUR>
 A;Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:g14972654; GSPDB:GN00164; TIGR:SP45
 A;Experimental source: strain TIGR4
 C;Genetics:
 A;Gene: SP1174
 C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 67.2%; Score 2797.5; DB 2; Length 819;
 Best Local Similarity 67.0%; Pred. No. 5.6e-145;
 Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;

QY 1 SYELGLYQARTKVENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYV 59
 Db 21 SYELGRYQAGQDKKESNRVAVIDGQAGQKAEHLTPDEVSKREGINAEQIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIIESELLMKDPNYKLKDEDIVNEVKGYYVVKDGYVYVYKLD 119
 Db 81 TSHGDHYHYNGKVPYDAIIESELLMKDPNYKLKDEDIVNEVKGYYVVKDGYVYVYKLD 140
 QY 120 AAHADNVRTKEEINRQKQESHQREGTTPNDGAVALARSGRYTTDDGYIFNASDIIED 179
 Db 141 AAHADNVRTKEEINRQKQESHQREGTTPNDGAVALARSGRYTTDDGYIFNASDIIED 197
 QY 180 TGDYIVPHGDHYHYIPKNELSAELAAEAFUSGRGNLSNSTRYRRONSNDTSTRNWWP 239
 Db 198 TGDYIVPHGDHYHYIPKNELSAELAAEAFUSGRGNLSNSTRYRRONSNDTSTRNWWP 246
 QY 240 VSNPQTNTNTSNNTNSQASQSDNDISLLKQYKLPQSRHVSDDGLVDPQAITSRT 299
 Db 247 NPAQFRLSENHLTVTPYHQ-NQGENISSLLRELKPLSERHVSDDGLVDPQAITSRT 305
 QY 300 TARGVAVPHGDHYHYIPYQSMGELEERARIIPLYRSHNWPDSPRPEQSPQTPPEPSP 359

306 TARGVAVPHGNHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPSPPEPSPQTPPEPSP 365
 360 GPOPAENLKIDNSLSVQVLRVGVGVFEKGIYRVFAKDPSETVKNLESKSKOE 419
 366 SPOPAENFID--EKLVEAVRVGVGVFEENGVSRYIPAKDLSAETAAGIDSKLAKOE 423
 420 SVSHTLTAKENAVPRDQEPYDKAYNLLTEAHKALFKNKGRNSDFQALDKLLERLNDEST 479
 424 SLSHKLTGKTDLPSDSRDFYNKAYDILLARIHQDLDLONKGRQVDFEALDNLLEBLKDVSS 483
 480 NKEKLVDDLLAFIAPITHPERLCKNSQIETDEVRIAQLADKYTSDGYIFDEHDIIS 539
 484 DKVKLVEDIIAFIAPIRHPERLCKNSQIETDEVRIAQLADKYTSDGYIFDEHDIIS 543
 540 DEGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDSAAAI 599
 544 DEGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDSAAAI 603
 600 YNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPKOHYHNIKFAWFDHHTYKAPNGVYLE 659
 604 YNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWDEGLYEAPKGYILE 663
 660 DLFAITIKYVVEHPDERPHSDNGMGNASEHVLKKGKDHSEDNKNPKFADEE-----PVEET 713
 664 DLLATVYKYVVEHPDERPHSDNGMGNASEHVLKKGKDHSEDNKNPKFADEE-----PVEET 723
 714 -----PAPEVPQVETEKVEAQKRAEVLILAKVTDSSLKANA 750
 724 PREKPOSEKPEPKTEPEEPSEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 783
 751 TETLAGRNNTLQIMDNNSIMAEAKLLALLKGS 785
 784 KETLTGLKNLLFGTDNNTIMAEAKLLALLKGS 818

RESULT 5
 G95115
 conserved hypothetical protein sp1003 [imported] - Streptococcus pneumoniae (strain TIGR
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
 neon, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; PMID:21357209; PMID:11463916
 A:Accession: G95115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:g14972476; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: sp1003
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 66.6%; Score 2772; DB 2; Length 839;
 Best Local Similarity 65.0%; Pred. No. 1.4e-143;
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATQKTEHLTPDEVSKREGINAEQIVIKITDQGYV 59
 21 SYELGRHQAGQVKESNRVSYIDGQAGQKAEHLTPDEVSKREGINAEQIVIKITDQGYV 80
 60 TSHGDDHYHNGKVPYDAIISBELLMKDPNYKLKDEIDVNEVKGGVYIKVDGKYVYLLKD 119
 81 TSHGDDHYHNGKVPYDAIISBELLMKDPNYQLKDSIDVNEIKGGVYIKVDGKYVYLLKD 140
 120 AAHADNVRTKEEINRQKQESQHSRREGTTPNDGAVALARQGRYTTDDGYIFNADSLIED 179
 141 AAHADNVRTKEEINRQKQESQHSRREGTTPNDGAVALARQGRYTTDDGYIFNADSLIED 198

180 TGDAYIVPHGDHYHYIPKNELSASELAAAEAFLSRGNSLNSRTYRQNSDNTSRTNWVP 239
 199 TGDAYIVPHGDHYHYIPKNELSASELAAAEAFYVNG-----KQSRPSSSSSYNA 247
 240 SVSNPTGTTNTSNNNTNSQASQSDNDISLLKQLYKLPLOSBBHVESDGLVDPQAQITSR 299
 248 NPAQPRLSNNHLLTVTPYHQ-NQGENISLLRELYAKPLSERHVSDDGLVDPQAQITSR 306
 300 TARGVAVPHGNHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPSPPEPSPQTPPEPSP 359
 307 TARGVAVPHGNHVFIPYQMSSELEKRIARIIPLRVRSNHWVDPSPPEPSPQTPPEPSP 366
 360 GPOPAENLK-IDSN-----SSLSVQVLRVGVGVFEKGIYRVFAKDPSETVKNLESK 415
 367 SPOPAENFID--EKLVEAVRVGVGVFEENGVSRYIPAKDLSAETAAGIDSKL 426
 416 SKESVSHHTLTAKENAVPRDQEPYDKAYNLLTEAHKALFKNKGRNSDFQALDKLLERLN 475
 427 AKOESLSHKLGAKTDLPSDSRDFYNKAYDILLARIHQDLDLONKGRQVDFEALDNLLEBLK 486
 476 DESNKEKLVDDLLAFIAPITHPERLCKNSQIETDEVRIAQLADKYTSDGYIFDEH 535
 487 DVPSDKVKLVDDLLAFIAPIRHPERLCKNSQIETDEVRIAQLADKYTSDGYIFDEH 546
 536 DIISDEGDVAVTPHMGSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDS 595
 547 DIISDEGDVAVTPHMGSHWIGKDSLSDEKVAQAAYTKEKGIILPPSPDADVKANPTGDS 606
 596 AAATYNRVKGKRIPLVRLPYMVEHTVEKNGNLIIPKOHYHNIKFAWFDHHTYKAPNG 655
 607 AAATYNRVKAARKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWDEGLYEAPK 666
 656 YTLRDLFATIKYVVEHPDERPHSDNGMGNASEHVLKKGKDHSEDNKNPKFADEE-----KOHSE 697
 667 YTLRDLFATIKYVVEHPDERPHSDNGMGNASEHVLKKGKDHSEDNKNPKFADEE-----KOHSE 726
 698 -----DNKNPKFADEEPEVETPAPEVPOVETEKVEAQKRAEVLILAKVTDSSLKANA 738
 727 ESDEKENHAGLNPSADNLKPSDTDETEPEAEADTDEAEIPQVENSVINAKIADAEALL 786
 739 AKYVTDSSLKANAETLAGRNNTLQIMDNNSIMAEAKLLALLKGSNPSPV 790
 787 EKVTDPSIRQAMETLTGLKSSLLGLTKDNNNTISAEVDSLLALLKESQAPAPI 838

RESULT 6
 C97985
 hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: C97985
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; PMID:21429245; PMID:11544234
 A:Accession: C97985
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-853 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:g15458514; GSPDB:GN00174
 C:Genetics:
 A:Gene: phtD
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 65.6%; Score 2733; DB 2; Length 853;
 Best Local Similarity 63.5%; Pred. No. 2e-141;
 Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATQKTEHLTPDEVSKREGINAEQIVIKITDQGYV 59

Db 21 SYELGRHQAGQVKESNRVIDGQAGQKAENLTPDEVSKREGINAEQIVIKITDQGVY 80
 QY TSHGDHYHYNGKVPYDAI ISEELLMKDPNYKLKDEDIVNEVKGYYIVKDGKYYVYLKD 119
 Db 81 TSHGDHYHYNGKVPYDAI ISEELLMKDPNYQLKSDIIVNEIKGGYIVKDGKYYVYLKD 140
 QY 120 AAHADNVRTKEEINRQKQSHREHGGTTPRNDGAVALARSOGRYTTDDGYIFNASDIIED 179
 Db 141 AAHADNVRTKEEINRQKQSHREHGGTTPRNDGAVALARSOGRYTTDDGYIFNASDIIED 197
 QY 180 TGDVAVPHGDHYHYIPKVELSASELAABAFPLSGRGNLSNRTVRRONSNTSRTNWVP 239
 Db 198 TGDVAVPHGDHYHYIPKVELSASELAABAFPLSGRGNLSNRTVRRONSNTSRTNWVP 239
 QY 240 SVSNPGTNTNTNNTSNTNSQASOSNDIDSLKOLYKLPISORHVESDGLVFPDPAQITSR 246
 Db 247 NPAQPRLSNENLTVTYTQ-NOGENISSULRELYAKPLSERHVESDGLVFPDPAQITSR 305
 QY 300 TARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPRPS 359
 Db 306 TANGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPRPS 365
 QY 360 GPQAPNLK-IDSN---SSLVSLQVRKVGYYVEEKGIGRYVPAKOLPSETVKNLESKL 415
 Db 366 SPQAPNPQAPNPIDEKLVKAVRKVGYYVEEKGIGRYVPAKOLPSETVKNLESKL 415
 QY 416 SKQESVSHLTAKKENVAPRQDFYDKAYNLLTEAHKALPKNGRNSDFQALDKLRLN 475
 Db 426 AKQESLSHKLGAKTDLPSDRFYNKAYDLARIHQDLNDKGRQVDFEALDNLRLK 485
 QY 476 DESTNKEKLVDDLAFAPITHPERLGKNSQIEYTEDVRIQALADKYTTSDGYIFDEH 535
 Db 486 DVSSDKVLDVDDLAFAPITHPERLGKNSQIEYTEDVRIQALADKYTTSDGYIFDEH 545
 QY 536 DIISDGDVAVPHGMGSHWIGKSDSKDEKVAQAAYTEKGLPSPDADVKANPTGDS 595
 Db 546 DIISDGDVAVPHGMGSHWIGKSDSKDEKVAQAAYTEKGLPSPDADVKANPTGDS 595
 QY 596 AAIYNRVKGKRIPLVRPYMVEHVEKNGNLIIPKDHVHNKIFAFDHYTAPNG 655
 Db 606 AAIYNRVKAACKVPLDRPNLYQYVEKNGSLIIPHYDHNKIFEFWDFEGLYBAPKG 665
 QY 656 YLEDLFATIKYVYHEDRPHNSDGNWGNASEHVLGK-----DHSEDPNK----- 701
 Db 666 YSLEDLLATVKYVYHEDRPHNSDGNWGNASEHVLGK-----DHSEDPNK----- 701
 QY 702 -----NFKAD-----EHPVETPABPEVQVE 723
 Db 726 EEDKEHDEVSETHPESDEKENHVGINPSADNLYKPTDTEETEBAEDTTDEAIPQVE 785
 QY 724 TEKVEAQLKEAEVLLAKVTDGSLKANATETLAGLRNNLTLOIMDNNSIMAEAKLLALIK 783
 Db 786 HSVINAKIAEAEALLEKVTDSIRQNAVETTLGLKSSLLGLTKONNTISAEVDSLLALLK 845
 QY 784 GSNSSV 790
 Db 846 ESQTPPI 852

RESULT 7
 H95115

conserved hypothetical protein sp1004 [imported] - Streptococcus pneumoniae (strain TIGR
 C/Species: Streptococcus pneumoniae
 C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C/Accession: H95115
 R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
 Science 293, 498-506, 2001
 A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A/Reference number: A95000; MUID:21357209; PMID:11463916

A/Accession: H95115
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1039 <KUR>
 A/Cross-references: GB:AB005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:SP4
 A/Experimental source: strain TIGR4
 C/Genetics:
 A/Gene: SP1004

Query Match 29.9%; Score 1246; DB 2; Length 1039;
 Best Local Similarity 40.9%; Pred. No. 3.6e-60;
 Matches 296; Conservative 97; Mismatches 180; Indels 150; Gaps 18;

QY 1 SYELGLYQARTVKENNRVSYIDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGVY 60
 Db 22 AYALNQHSQENKNNRVSYVDGSSQSKSENLPDQVSQKQEQAGQIVIKITDQGVY 81
 QY 61 SHGDHYHYNGKVPYDAI ISEELLMKDPNYKLKDEDIVNEVKGYYIVKDGKYYVYLKDA 120
 Db 82 SHGDHYHYNGKVPYDAI ISEELLMKDPNYQLKSDIIVNEVKGYYIVKDGKYYVYLKDA 141
 QY 121 AAHADNVRTKEEINRQKQSHREHGGTTPRNDGAVALARSOGRYTTDDGYIFNASDIIED 180
 Db 142 AAHADNVRTKEEINRQKQSHREHGGTTPRNDGAVALARSOGRYTTDDGYIFNASDIIED 197
 QY 181 GDVAVPHGDHYHYIPKVELSASELAABAFPLSGRGNLSNRTVRRONSNTSRTNWVP 240
 Db 198 GNAYIVPHGDHYHYIPKVELSASELAABAFPLSGRGNLSNRTVRRONSNTSRTNWVP 248
 QY 241 VSNPGTNTNTNNTSNTNSQASOSNDIDSLKOLYKLPISORHVESDGLVFPDPAQITSR 300
 Db 249 -----NTQSAVAGSTSKPANSENLSLLKELYDPSAQRYSESGLVFPDPAQITSR 301
 QY 301 ARGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPRPS 360
 Db 302 PNGVAVPHGDHYHYIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPRPS 360
 QY 361 POPAPNLKIDSNSSLVSLQVRKVGYYVEEKGIGRYVPAKOLPSETVKNLESKLKQES 420
 Db 334 -----ISGTGSTVSTNAK-----PNEVSSLSGLSNPSS 363
 QY 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALPKNGRNSDFQALDKLRLNDESTN 480
 Db 364 -----LTSKELSSASDGYIFNPK-DIVEETATAYVRHG--DHFHYIPK-----SNQIG 410
 QY 481 KEKLVDDLAFAPITHPERLGKNSQIEYTEDVRIQALADKYTTSDGYIFDEHDIISD 540
 Db 411 QPTLPNLSLATPSP-SLPINPOTSEKHE-----EDGYGFDANRIIAE 452
 QY 541 EGDVAVPHGMGSHWIGKSDSKDEKVAQAAYTEKGLPSPDADVKANPTGDSAAIY 600
 Db 453 DESGFVMSHGDHNYFFKKDLTEEQIKAAQKHEE-----VTSINGLDSLSH 501
 QY 601 NR-----VKGEKRIPLVRPYMVEHVEKNGNLIIPKDHVHNKIFAFDHYTAPNG 644
 Db 502 EDDYPSNAKEMKDLKKIEEKIAGIMQYGVKRESIVVNEKKNALIIYPHGDHHDADP--- 558
 QY 645 PDDHYTAPNGYTTLEDLFATIKYVYHEDRPHNSDGNWGNASEHVLGKSDSKDEKVAQAAYTEKGLPSPDADVKANPTGDSAAIY 700
 Db 559 IDEH---KPVIG-----HSHSNYELFPKEGVAKEGKNKVTGTBELTNVNLKNSTFNN 611
 QY 701 KNF 703
 Db 612 QNF 614

RESULT 8

D97985
 hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
 C/Species: Streptococcus pneumoniae
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C/Accession: D97985
 R/Hobkins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Es

hypothetical 92.4K protein - Streptococcus agalactiae
 C:Species: Streptococcus agalactiae
 C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
 C:Accession: T46758
 R:Spellerberg, B.; Rodzinski, E.; Martin, S.; Weber-Heynenmann, J.; Schnitzler, N.; Luett
 Infect. Immun. 67, 871-878, 1999
 A:Title: lmb, a protein with similarities to the Lral adhesin family, mediates attachment
 A:Reference number: Z24091; MUID:99115568; PMID:9916102
 A:Accession: T46758
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-822 <SPE>
 A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624
 C:Experimental source: strain R268
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.4%; Score 934; DB 2; Length 822;
 Best Local Similarity 29.4%; Pred. No. 2.7e-43;
 Matches 271; Conservative 99; Mismatches 235; Indels 318; Gaps 32;
 1 SYELGLYQARTVKENNRVSVID--GKQATQKTENLTPDEVSKREGINAEQIVIKITDQGY 57
 22 AYALNQRSENKDNRRSVYDGSQSSQKSENLTDPQVQKGLQAEQIVIKITDQGYVT 81
 61 SHGDHYHYNGKVPYDAIIESEELMKDPNKLKDEDIVNEVKGGYVIKVDGKYVYLKDA 120
 82 SHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGGYIIVKVDGKYVYLKDA 141
 121 AHADNVRTKEINRQKQEHSGHREGGTRNDGAVALARSQGRYTTDGYIFNADITIEDT 180
 142 AHADNVRTKEINRQKQEHVKNDE---KVNNSNVAVARSQGRYTTDGYFNPADITIEDT 197
 181 GDVIVPHGDHYHYIPKNELSASELAAAEFLSGRNLNSRTYRRQNSDNTSRTNWVPS 240
 198 GNAYIVPHGGHYHYIPKSDLSASELAAAHAGLAKWQPSQVSSSTASDN-----248
 241 VSNPGTNTNNTSNTSQAQSDNDISLKLQYKLPVLSQHVESDGLVDPDAQITSR 300
 249 -----NQVAKGTSKPAKNSLENLQSLKELYDSPAQRYSESGLVDPDAQITSR 301
 301 ARGVAVPHGDHYHYIPYSQSELEERARIIPLYRSHWVDSRPEQSPQTPPEPSG 360
 302 PNGVAIPHGDHYHYIPYKLSALEEKIARRVP-----333
 361 POPAPNLKIDNSNLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKQES 420
 334 -----ISGTGSTVSTNAK-----PNEVSSLSGLSSNPSS 363
 421 VSHLTAKENVAPRQDFYDKAYNLLTEAHKALFKNKGRNSDFQALDKLLERLNDESTN 480
 364 -----LTSKELSSASDGIYFNPK-DIVBETATAYIVRHG--DHFHYIYK-----SNQIG 410
 481 KEKLVDDLLAFIAPITHPERLQKP-NSQIEYTEDVRIQAQADKYTTSDGYIFDEHDIIS 539
 411 QPTLPNNSLATPSP-----SLPINPGISHEKHE-----EDGYGFANDRIIA 451
 540 DEGDAYVTPHMGSHWIGKDSLDKEKVAQAAYTKEGILPPSPDADKANPTG-DSAAA 598
 452 EDESGFIMSHGNHNYFFKDLTEEQIKAAQKHLE-----VKTSHGLDLSLS 500
 599 IYNRVKG-----EKRIPLVRLPYM--EHTVEKGNLLI-PHKDHYHNIKEA 643
 501 HEQDYPGNAKEMKDLKKEIKIAGIMKQYGVKRESIVVNEKNAIYPHGDHHDADP--558
 644 WFDHDTYKAPNGYTTLEDLFATIKYVYEHDPDRPHSNDG---WGNASEHVLGKKHSDP 699
 559 -IDEH---RPVIGG---HSHSNYELFKPEGVAKKEGKNKYVTGEBLTNVNLLKNSTFN 610
 700 NKNF 703
 611 NQNF 614

Db 633 VTPIDKRAGMENWOIHSAEVQKALAEGRFAAPDGYIFDPDVLAKETFWKDGSP 692
Qy 721 QVE-----TEKVEAQKAEAEVLLAKVTDDSSLKANATETLAGLRNNLTQLQIMDNNSI 771
Db 693 RADGSSRLTINKSDLSAQEWQQAQELLAKKN---AGDATDT-----DXPEE 735
Qy 772 MAEAEKLLALLKGSNPPSSVSKEK 794
Db 736 KQQAQK---SNENQOPSEASKEE 755

RESULT 10
P97985
hypotheical protein pheE-truncation [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: P97985
E/Hoskins J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, P.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: P97985
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-182 <GR>
A/Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:gl5458517; GSPDB:GNO0174
C/Genetics:
A/Gene: pheE-truncation

Query Match	5.8%;	Score 243;	DB 2;	Length 182;
Best Local Similarity	31.6%;	Pred. No. 1.5e-06;		
Matches	61;	Conservative 16;	Mismatches 50;	Indels 66; Gaps 5;
QY	147	TPRNDGAVALARASQGRYTTDDGYIFNASDILIEDTGAYIVPGDHVHYIPKNELSASELA	206	
DB	4	TPNN-----GVSAYDDGVFNPNDIVRDTGDYIVRGGDHVHYIPKNSLN-----	48	
QY	207	AABAFLSGRGNLSNRYRRQNSDNTSRTNWVPSVNPCTNTNTNSNSTNSQASQSD	266	
DB	49	-----NPPSHSNTTEEVGSSSS	65	
QY	267	IDSLLKQLYKLPFSQRHVESDGLVFPDPAOITSTARGAVAPHGHDVHFIPYQSMGELEER	326	
DB	66	V-----LSNPSLHVHHEEDGHGFANDRII SEDSGFVIPGHDHNYIKV-QTKGYEAA	118	
QY	327	IARIIPLAYRSNH	339	
DB	119	LNKKIP-SLQSNY	130	

RESULT 11
Tt14004
C:Species: Dictyostelium discoideum
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: Tt14004
C:Accession: Tt14004
J. Biol. Chem. 273, 24654-24659, 1998
A:Title: Dictyostelium TRPA homologous to yeast Ssn6 is required for normal growth and
A:Reference number: Z17852; MUID:98406112; PMID:9733762
A:Accession: Tt14004
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1390 <SAI>
A:Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BA033143.1
C:Genetics:
A:Gene: trfA
A:Exons: 333/3; 364/3; 637/1
A:Introns: 333/3; 364/3; 637/1

			Query Match	4.6%; Score 192.5; DB 2; Length 1390;
			Best Local Similarity	18.1%; Pred. No. 0.016;
			Matches	147; Conservative 131; Mismatches 291; Indels 245; Gaps 35
QY	15	NNRVSVLDGKQATOKTENLTPEVKSRGINAEQIVIKITDGYVTSHGDHYHYNKGVP	74	
Db	690	NNNNNINNNMNNNNNNHHNMQMYYNN-----NNSNNNNNNNNHHNN-----	736	
QY	75	YDALISELLMKDPNYKLKDIDLVNEVGKVYIKV-----DGKYVVYLKDAAHADNVRT	128	
Db	737	-----DNVSKNDVDLRRYKGLIREKTSPNGDR-----DNRDNTRD	775	
QY	129	KEEINRQOEHSQHREGGTTPR-----NDCAVALA-----RSQGRTYTDDG	168	
Db	776	NRD-NRDGRDGRDNRDSRDDRIQEYTRYENNNNNNNSSINNNNNNNNNNYNNWN	834	
QY	169	YIFNASDIIEDTGDAYIYP-----HGHHYHIYIPNELSASELAALAAEFLSGRGLSNSTRY	225	
Db	835	NNNNNNNNNNGLRATPLPSHNDRRSYERDK-----BRINNNNNNNNNNMM	885	
QY	226	RONSNTSRTKWVPSV-----SNPGTTINTTSNNSTNSSASOSNDID-SLLKQL	274	
Db	886	NNNNNNNNNNNISISFEHQNNRTSPFFENNNNNNNNNN-----NNNINIINIGQR	941	
QY	275	YKLPLSORHVESDGIADF-PAQITSRT-ARGVAVPHGDHYHFIPVSOMSELEERIARIIP	332	
Db	942	ALSPOSSHQKRREILDESDINERSKTRFSPI-----VKEABEKRETIVY	988	
QY	333	LAYSNNHWPDSPQPSPQPTPPSPGPAPNLKI-DNSSSLVSOLVRKGVGGYVFEE	391	
Db	989	DKERS-----PPTIIETKPQEVKVTDESSLVEKVDXE-----NEK	1027	
QY	392	KGISRYVPAKDLPSETVNLESKLSKOESVSHTLTAKENVAPRDQFYDYKANLLTEAH	451	
Db	1028	ESPSSSSSKIEKETBKBEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE	1078	
QY	452	KALFXNKGRNSDFOALKLLER-----LNDESTNKEKLVDDLAFAPITHPERLGKPN	505	
Db	1079	KDVENEK---SVEKSSEPKVESTTTTTNDEDGE-----LSEPTTTTKCD	1123	
QY	506	SQIETYTBDEVRIAQLADKYTSDGYIFDEHDLIISDEGDAYVTVMGHSHWIGDKSLSDKE	565	
Db	1124	DSSKLPTDEKKLSVSPPTTAIVEQ-----SDETKELE	1156	
QY	566	KVAQAQATKEGILPPSPDADVXANPGDGAIA-IYNRVKGEKRIPLVRLPYMVEHTVTV	624	
Db	1157	MOTKEDSEKEX-----KSTTTTTAAAGESVKPIDEEKSPITT-----TTTT	1198	
QY	625	KGNGLIIP-HKHVHNKFAWFDDHT-----YKA PNGYILEDLFATI KYVYEHDP	673	
Db	1199	TNITTVETHKDESSKN-----DTTTTTTTTTTTKSAKSPNSPTRS-----D	1242	
QY	674	E-RPHSNDGWGNASHVLGKDHSED-----PNKNFKADEEPVEETPA-----EPBVQV	722	
Db	1243	EVVEPHQ-----DASQEBEINKRKLEDILTSTPSKRLKPDSTSSATTASTPSBPESP-L	1296	
QY	723	ETEKVBQALKEAVILLAKVTDSSLKANAFETTAG	756	
Db	1297	KCNPNVGETLSPFIKDSKSSSSSSSSSSSSSSSTNG	1330	

RESULT 12
A45555
glutamate rich protein - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C/Accession: A45555; S27831
R/Borre, M.B.; Dziedzic, M.; Hough, B.; Petersen, E.; Rieneck, K.; Riley, E.; Meis, J.F.;
J.
Mol. Biochem. Parasitol. 49, 119-131, 1991
A/Title: Primary structure and localization of a conserved immunogenic Plasmodium falciparum
rate life cycle.
A/Reference number: A45555; MUID:92131041; PMID:1775153

A;Accession: A4555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1271 <BOR>
A;Cross-references: EMBL:M59706; NID:g160311; PID:g160312
A;Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBI:P:77802)

Query Match 4.2%; Score 176.5; DB 2; Length 1271; Gaps 43;
Best Local Similarity 19.1%; Pred. No. 0.11;
Matches 170; Conservative 140; Mismatches 347; Indels 231;

QY 7 QYATVKNRVSVIDGKQATKTNLT-----DEVSKREGINAEQIVIKITDQGY---- 58
DB 373 HNINVLQNNINNHQLEPQEPKNIPEPKNIDSEIILPENVEEIIIDVPSKHSNHE 432
QY 59 ----VTGSHGHHYNGKVPYDAIISELLMKDPNKLKOBIDIVNEVKGKVIKVDGKYY 114
DB 433 TFEETSESEHEEAVSEKNAHEITVEHEETVSQESNPEKADND-----GNV----- 477

QY 115 VYLKDAHAADN-VTKKEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNA 173
DB 478 ----SONSNELNENEFVESEKSEHE-----AAEESSELEEGH---H 513

QY 174 SDII-----EDTGAYIVPHGDHVIYIPKNELSAELAAAEFLSGRNLNRSRYRQ- 227
DB 514 EEIYPEQNNESGESKLV-----DNEGGEFAHNFSS---EVSNSLNENEF 560

QY 228 -NSDNTSRTNWVPSNFGTTNTNTNNSNTNSQASQNDISLLKQYKLPQSRHVES 286
DB 561 VESDK-----SVTEPAEHEEVSESNPEAENESSIEEAHQE--EIVPEQNDDES 610

QY 287 ---DGLV-----FDPAGITS-----RTARGVAVPHGDHVIYIPYSOMSE 322
DB 611 GESGLVDNEEGDFEENHEEPDQNDSELSENLVESEKVSSEPA-AEHVEIVSEKVSSE 669

QY 323 LEERTARIIPLYRNHNWVPSRPPQPPQPTPEPS---PGPQAPNL-----KIDSNSSL 375
DB 670 PAEHV-EIVSEKSTSE---PAHVESVSEQSNNESEKKDGPVPSKPEEIEKVDQPKI 725

QY 376 VSQVRKVGEGYVF-----EEKIGSYVPAKLPSETVK---MLESKLSQESVSHLTAK 428
DB 726 VD-LQIIEPNFVDSQNPQEPVPSFKIEKVPSEENKASVDPEVKEKENVSEVVEEK 783

QY 429 K-----ENVAPDQEFYDKAYNLLTEAHK-----LFXNKGNSDFQALD 468
DB 784 QNSQESVEEIPVNEDEFEDVHTQDLQDHTVDPEIVEVEEIPSELHENEVAHPEIVEIE 843

QY 469 KLLERLNDESTNKEKLVDDLLAFLA-PITHPERLAKPNSQIERYTDE-VRIAQLADKYTT 526
DB 844 EVFPEPNQNNNEQNEEDDKSAHIQHEIIVEVEEILPEDDKNEKVEHEIIVEVEEILPEDKN 903

QY 527 SDGYIFDEHDIISDEGDVYTPHMGSHWIGKDSLDKEKVAQAQYATKKGILPSPDAD 586
DB 904 EKG-----QHEIVEVE---EILPE-----DDKNEKVEHEIIVEVEEILP--EDKN 942

QY 587 VKANPTGDSAAAIYNEVKEKRIPLVPLPMYVEH-TVEVKNGNLIIPKDHVHNKFAWF 645
DB 943 ERGQHEIIVEVEEILPEDKNEK-----VEHEIIVEVEE---ILPEDKNEKG----- 983

QY 646 DQHTYKAPNGYTTLEDLFATIKYVVEHPDE-RPHSNDGWGNAS-----EHLVGKKDSE--- 697
DB 984 -QHEIIVEVEEILPEDKNEKQHEIIVEVEEILPEDKNEKQHEIIVEVEEILPEDKNEKGQ 1042

QY 698 -----DPNKPKADEEPEETPAEPVQVTEKVEAQLKAEVLLAKVITDSSLK 747
DB 1043 HEIIVEVEEILPEDKNEKQHEIIVEVEEILPEDKNEKQHEIIVEVEEILPEIVE----- 1095

QY 748 ANATETLAGLRNLIQLIQIMNNSISMAEAKLLALLKGSNPPSVSSEKI 795
DB 1096 --IEEVPSQNTNNEIE---TIKPEKK-----NEFSVEEKAI 1128

D90011
FmtB protein [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: D90011
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: D90011
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2481 <KUR>
A;Cross-references: GB:BA000018; PID:g13701961; PIDN:BA043253.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: fmtB (mrp)

Query Match 4.1%; Score 172; DB 2; Length 2481;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 169; Conservative 140; Mismatches 353; Indels 210; Gaps 38;

QY 8 QARTVKNRVSVIDGKQATKTNLT-----ENLTFDEVSKREGINAEQIVIKITDQGYVTS 61
DB 969 QAAATKKSADKAEIAQKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1018

QY 62 HGD-----HHVYNGKVPYDAIISELLMKDPNKLKDEIDIVNEVKGKVIKVDGKYY 115
DB 1019 NADIDNATANTVDNAKTITNEATTA--AITPDANVKAQAIAD-----KVAQAQ--- 1066

QY 116 YLKDAHAADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNA 159
DB 1067 --ETALDANNSTTEEEAAKQVQTEKTA--DAIDAHAHSNVEVEAAKNABIAKIE 1120

QY 160 ---QGRYTTDDGYIFNASDIIEDTGAYIVPHGDHVIYIPKNELSAELAAAEFLSGRNL 217
DB 1121 AIQPATTTKD---NAQAIATKANERKTA-----IAQTQDITAEETAAANA-----D 1164

QY 218 LNSRTYERQNSDNTSRTNWVPSNFGTTNTNTNNSNTNSQASQNDISLLKQYKLP 277
DB 1165 VDNVAVTQANSIEAANSQNDVDQAKTTGETSID-QVTFVKNKATARNEITALNN--KL 1221

QY 278 PLSQRHVESDGLVFPDPAQITRTARGVAVPHGDHVIYIPYSOMSELEERARIIPRYRS 337
DB 1222 QEIQATPDATDEKQAADAEANTENGKA-----NQAIASAATTNAQVDEAKANA 1269

QY 338 NHWVPSRPPQSPQPTPEPSGPPQA-----PNLKIDSNSSVLSQVLRKVGEG--- 386
DB 1270 EAAINAVTPKVVKQAADKQIDQLOATQTNVINDQATNNEKEKAIQOLATAVTDKNN 1329

QY 387 ---YVFEKIGSYVPAKLPSETVKNLES-----KLSQESVSHLTILTKAKENVAPRQDEFY 440
DB 1330 ITAAATDDNGVDLT---AKDAGKNSIGSTQTPATVKNKAKNEVDQAVTTQNAI----- 1378

QY 441 DKAYNLLTEAHKALFXNKGNSDFQALDKLLER-INDESTN-----KEKLVDDLLAFLAP 494
DB 1379 DNTTGATTEE-----KNAAKDLVLKAKKAYQDILNAQTNTDVTQIKDQAVADIQGITAD 1433

QY 495 ITHPERLGPNSQIERYTDEVTIAQLADKYTTSDGYIFDEHDIISDEGDVYTPHMGSH 554
DB 1434 TTIKO-VAKDELATKANEOKALIAQTADATT-----BEKQANQVQDQLT--QGNQN 1483

QY 555 WIGKDSLSDEKVAQAQYATKKGILPSPDADVKANPTGDSAAAIYNRVKEKRIPLVRL 614
DB 1484 IENAQSIDDVNTAKNAI---QAIDPIQASTVDVKNARAELLTEMQNKI-----TEI 1532

QY 615 PYMVEHTEVKNGLIIP-----HKDHYHNKFAWF--DDHTYKAPNGYTTLEDLFATIKY 668
DB 1533 LNNNETTNEEK-GNDIGPVRAAYEEGLNINNAATTTGDTVTTAKTAVQKVVQQLHA----- 1586

QY 669 VEHPDERPHSNDGWGNASEHVLGKKDSEHPNKNFKADEE-PVEETP--AEPEVPQVETE 725

Db 1587 --NPVKKP-----AGKKELQ-----AAADKKTQLEQTPNASEQINDAKOE 1626
QY 726 KYEAOLKEA-----EVLAKVTDSSKANATETLAGNNLTLOIMD----- 767
Db 1627 -VDTELOAKTNVDSSTNEYVDNAVKEGAKINAVKTFSEYKQDALAKIEDYNAKVNE 1685
QY 768 -----NNSINAEAEKLLALLKGSNPSSVSK 792
Db 1686 ADNSNASTSSEAEAKQKLAELKQTADQNVNQ 1717
RESULT 14
S56271
hypothetical protein YFR016c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
C:Accession: S56271
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasano
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*
A:Reference number: S56186
A:Accession: S56271
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1233 <MUR>
A:Cross-references: EMBL:D50617; NID:G836685; PIDN:BAA09255.1; PID:dl009896; PID:G836771
C:Genetics:
A:Map position: 6R

Query Match 4.1%; Score 171; DB 2; Length 1233;
Best Local Similarity 19.8%; Pred. No. 0.21;
Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;
QY 12 VKNRVSIDYDQATOKTENLTDPDEVSKEGINAEQIV-----KITD-QGYVTS 61
Db 151 IKETSTNNVAG-----TENVPP--IKESTGIEVGSNPTTRKKNKKKTKTNRGRNSS 202
QY 62 HGDHYHYNGKVPYDAII--SELLMKDPNKLKDEDI-----VNEVGKGYVTKVD 110
Db 203 NPADTTDLKQSTLDSILVIEEVLQEDGS-----KNEDIKNIVODEPNNVNEKMDIRTRNE 259
QY 111 GKYYVYLKDAHADNV-----RTKEEINRQKQHSQHREGG--TPRNDGAVALARQ--- 160
Db 260 SSDTKTFDIDVPNDVNDVETSSKGENNINEEKAETLPRNEILNVNEGNAAFKQLEP 319
QY 161 -GRYTTDDGYIFNADIIEDTGAYIVPHGDHYHYIPKNELASAEAAAEAFSLGRNLS 219
Db 320 HGLEAGDE-----NGQASTKDVESESLTKNGFNKFNESKHLKAGEKQOTE---SDRDGIS 372
QY 220 NSRTYRRQNSDNTSRTNWPVSVSNP-----GTTNTNTNNSNTNSQASQSNDI----- 267
Db 373 PSVLAKNQKETEIGKEDHVFPEQKDEKCKRKLSSVNNHNNHNFNAAGSDIIPPETE 432
QY 268 -----DSLKLQLYKLPLSRHV-----ESDGLVDPQAQ---ITSR 299
Db 433 RETYDDTGMPTKRIISDNENKLOHGTNDISVEVEKEEEEEEEENSTSKVKNENTGE 492
QY 300 -----TARGVAVPHGDHYHFPYSQMSLEERARIIPLYRSN----- 338
Db 493 QEAVRNNEVSGTEESTSGKEEIMGD-----EQSEBAGEK-SSIIIEGANSKAKIS 544
QY 339 --HWVPSRPEOPSPQPTPEPSPGPQAPNKLKIDSSSLVSLQVRKVGSGYVFEKIGSR 396
Db 545 KNLVLDEDEAEAPTQENKTEVVGEIDIDAPRD-----VELVEAVENKIIPEDLEVA 599
QY 397 -----YVFAKQIDPSETVKNLSEKLSQESVSHLTAKENAVPR-----DOEFYDK 442
Db 600 EDQEGEQVLKDEPVRAMKDDKIAMRGAESISDDMKKKQEGTAEKLSNEKAKKEVDETARE 659
QY 443 AYNLLTEAHKALFXNK-----GRNSDFOALDKLLERLNDESTNKKELVDOLLAPLAPI 495

Db 660 AEGVEVEKSTPESPKVVKRCTSGRPEDLQINERDPVLKEDVRVDEDKPEIATTIEN 719
QY 496 THPERLGKPNQIETEDERVIAQLADKYTTSDGYIFDEHD-----IISDEGDAYVTPHM 550
Db 720 SEEDPKSORVQISTEQATTKDKMDGVGSTTS---FKEEKKPKFEITQEGDKITGKDT 776
QY 551 GHSHWIGKDSLDKKEKAAQAYTKGKILPPS-----PDADVKANFTGDSAAAIYNR 602
Db 777 NHEHGEATEAASENSK-ASDVGTAETKYEIPSSSVKVKDTEEDAEV-----NSEKTEFIK 830
QY 603 VKGEKRIPLVRLPYMVVHTVVKNGNLI-----PHKHYNHNIKAFMFDHTY 650
Db 831 VKAE--LENLDAPKAEVTAELNKENEDVEVDTEEDAEVENSEKTEFIKVK---AELGNL 885
QY 651 KAPN--GYTLE-----DLFATIKYYVE---HPDERPHSNDGNGNASEHVLCKD--- 694
Db 886 DAPKAEVTAELNKENEDVEVAATSKEDIEFKCSEPAETPIED---GTCTEAEVSKKDAE 942
QY 695 --HSDPNKNKADDEPVEETPAPPEVQVE-----TEKVE-AQLKEAEVLLAKVTDSSL 746
Db 943 AVTKEDENMNSKIAEALKDVTGQEQEIDIDNISDEFQRTVELPELEKQIDKNKGEDKEL 1002
QY 747 KANATETLAGI-----RNNLTLOIMDNNSINAEAEKLLALLKGSNPSSVSEKIN 796
Db 1003 EVEETEKETSLPDLVVENIT---EEKNEIKQEEEBE-VSOLDNFNETESISKEAPN 1053

RESULT 15

D84900
hypothetical protein At2g46240 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84900
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.J.
euss, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84900
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1043 <STO>
A:Cross-references: GB:AE002093; NID:g3702325; PIDN:AAC62882.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g46240
A:Map position: 2

Query Match 4.0%; Score 168.5; DB 2; Length 1043;
Best Local Similarity 18.5%; Pred. No. 0.22;
Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 38;

QY 40 KREGINAEQIVIKITDQGYVTSBGDHYHYNGKV---PYDAIISELLMKDPNKLKDED 96
Db 271 KRDDVEASE-----SSNEDKKQNGKGTVEYFPDIDSMIKSL----- 306
QY 97 IVNEVGKGYVIKVDGKYVYVYLKDAHADNVKKEI-----NRQKQHSQH 142
Db 307 ----IQG-----QDVKEAQNKKEEPQVYPYFIPWIPSGKAKDVEASS 348
QY 143 REGGTP-----RNDGAVALARSQGYTTDDGYIFNADIIEDTGAYIVPHGD 190
Db 349 KESSNEGRNLESCPSDLHRNEQGITQAKGK-----EGNFEKNVLSDAEKKSVINIPVAN 403
QY 191 HV-----HVIPKNELASAEAAAEAFSLGR----- 215
Db 404 HLOEPRNIPVKLSEHNLFPAPTEPTKRIAKNEPVKTKKEQSSSSSEASKLPVCLRVDP 463
QY 216 -----GNLSNRTYRRQNSDNT-----SRTNWPVSVNPGTNTNTS---NN 254
Db 464 PKERNCGSKSVSHPKRMEKSKETKIAAPLSSKKAESRT--VPEACNVKCEDANAEMKMAE 521
QY 255 SNTNSQASQSNIDSLKLQLYKLPLSRHVESDGLVDPQAQITSRRTARGVAVPHGDHYHF 314

Db 522 GSNALRTEKGVES-----NSMQESNGEIKICEAKENRQP----- 561

Qy 315 IPIYOMSELEERLARIIPLYRS---NHWVDSRPEQPSPQTPPEPSPG----- 360

Db 562 ---AKSFTTEEARIIQSMVGVYDVERWEPFIKKLEIA---TVREQMGVKKRIEALFA 615

Qy 361 --PQAPNLKIDNSSILVQVRKGVYFEEKIGISRYVFAKDLPSFTVKNLESKLS-- 416

Db 616 STDQHIEKEIVNNGELVNMILLKLD---AVEGLHPSIREFRKALATE--LSSIQDKLDSL 671

Qy 417 KQBSVSHTLTAKKENVAPROEFYDKAVNL---LTRAHKAFLXNKGNSDFQALDKLLE 472

Db 672 KNSCASAKEAVQEVELKSP--SDSPVNLBHSQLTENKVV----- 712

Qy 473 RLNDSTNKEKLVDDLLAFLAPIHPERLKGPNQIEYTEDVRIAQADKYTSDGVIF 532

Db 713 ---SDTNLEKVL---LSPEEH-----PMSVLNRTDEK---QAESAAATEEGY-- 751

Qy 533 DEHDIISDEGDAYVTPHMGHSHWIGKDSLSKKEKVAQAAYT-----KEGILPPSPDA 585

Db 752 ---PSADGNGMTVNVEENKAMVLESLEPIELPQVBETETNSIRDPENASEVSEA 791

Qy 586 DVKANPTGDSAAIYVRYKGEKRI-----PLVRLPYMVEHT-----VEV--- 624

Db 792 ---PSADGNGMTVNVEENKAMVLESLEPIELPQVBETETNSIRDPENASEVSEA 846

Qy 625 -----KNGNLIIPHKD---HYHNKIFAFDHDH-----YKAPNGYTLDFEA--- 663

Db 847 ETNSENENRKGEDIVLHSEKNVSELPVGVIDEETQPLSQDSSSYTREGNMTAMD 906

Qy 664 -TIKYVVEHPDERPHSNDSNGWGNASHVLGKKHSDPNKNFKADEPVE-----ETPAE 716

Db 907 KTASQETEVDSHSPNSKIGIGQOTSEPODEKEQS--PETEVIVKQPLETEVILNEQAP 964

Qy 717 PVPQVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLNNLTQLIMNNSIMAEAE 776

Db 965 PEITEPGISKETKLMENORFETMETILVKAQREQ--LEVISKLSRVKSEKKLSHKK 1022

Qy 777 KLLALLKGSNPSSVS 791

Db 1023 KTQIRRRASKPWSVS 1037

RESULT 16

G89957

hypothetical protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: G89957

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89957

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KUR>

A:Cross-references: GB:BA000018; PID:g13701526; PIDN:BAB42820.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA1552

Query Match

Best Local Similarity 4.0%; Score 165; DB 2; Length 891;

Matches 177; Conservative 140; Mismatches 323; Indels 292; Gaps 48;

Qy 7 YQARTVKNRVSVDKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVTHGDHY 66

Db 36 HQAQAENTNTSDKISENNQNNATTQOPKDTNQTQ--PATQPV--ITAKNYPAA----- 86

Qy 67 HYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGYYVI---KVDGKYVYV--LKDA 121

Db 87 -----DESLKDAI--KDPALENKEHDIGPREQVNFQLLDKNNETQYHFFSFKDPA 135

Qy 122 HADNVRTKEI-----NROK-----QEHSHQREGTGRNDG 152

Db 136 DVYVYTKKAEVELDINTASTWKVFEVYENQKLPVRLVSYSPVPEDHAYIR---FPVSDG 192

Qy 153 AVALARSQGRYTTDDGY-----IFNASDIIE-DTGDAIVPHGDHYHYPK 197

Db 193 TQEL-KIVSTQIDDEETNYDTKLVPKPIYNDPSLVKSDTNDAAVVT----- 240

Qy 198 NELSASLAAABAPLSGRGNLSNRTYRRQNSDNTSRNWV-----SVSNPG----- 245

Db 241 NDQSSSDAS-----NQNTNT-SNQNTSTNNANNOPOATTNNMQPAQKSSANA 289

Qy 246 -----TTNTNTSNNTNSQASQSDNIDSLKQLYKLPLSQRHVESDGLVFD----- 292

Db 290 DQASSQPAHETNSNGNTNDKTNESNQSD-----VNOQYPPADESLQDAIKN 336

Qy 293 PAQITSTRTARGVAVPHGDHYHIFIPYSQMSLEERARIIPLYRSNHNWVDSRPEQSPQ 352

Db 337 PAIDKEHT-----ADNWRPIDQMKDKGER-----QFYHYA----- 369

Qy 353 PTPPSP-----PQGPAPNLK-----IDNSLSLVQLYR-KVGEGVVPEEKGIS 395

Db 370 STVEPATIVFTKTGVTIELGHTASTWKKFVEYEGDKLPVELVSYSDSKDYAIRFPVS 429

Qy 396 RYVFAKOLPSTVKNLESKLSQBSVSHTLTAKKENVAPRQDEFYD-KAYN---LLTBAH 451

Db 430 N--GTRDV--KNVSEIEGENIHEDYDTLMVFAQPIINNPDYVDEETYNLQKLLAPVH 485

Qy 452 KALFKNKGRNSDFQALDKLERLNDE-STNKEKLVDDLLAFLA-----PIHPERLGPEN 505

Db 486 KA---KTLERQVVELEKLEKLPKRYKAEYKKLDQTRVELADQVKSATVEFENVPTN 541

Qy 506 SQIETEDEVRIAQADKYTTS--DGYIFDEHDIISD--EGDAYVT--PHMGHSHWIGKDS 560

Db 542 DQITDQSAHFVVFSEENSESVMDGFV--EHPFYATILNGQYVVMKTKDDSYN--KDL 597

Qy 561 LSDKEKVAQAAYTEKG---ILPPSPDADVKANPTGDSAAIYVNRVKGKRIPLVRLDY 616

Db 598 IVEGHRVTVSKDPKNNSTLIFPIPD-----KAVYNAL---VKVWVANIGY 642

Qy 617 MVEHTEVKNGLNLIIPHKDHYHNI-----KPAWFD--DHTYKAENGYTLEDL 661

Db 643 EQGVHVRILNQDINTKDDDTSQNTTSEPLNVQTGQEGKVADTTVAENSTATNPKDASDK 702

Qy 662 FATIKYVVEHPDERPHSN--DQWGNASEHVLGKKHSDPNKNFKADEPVEETPABPE 718

Db 703 ADVIE-----PESDVVKDADNNDKDVQHDVHLSMDSDNNHFD----- 741

Qy 719 VPQVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLNNLTQLIMNNSIMAEAE 763

Db 742 -----KYDLKEMDTQIAKDTDRNV-----DNSVGMSSNVDTDKDSNKNKDKVIOLA 787

Qy 764 QIMDNNSIMAEAEKLLALLKG-SNPSSVSKEK 794

Db 788 HIADKNHHTGKAALDQVVKQVNNNTDKVTDKK 819

RESULT 17

A44361

amiloride-sensitive sodium channel Apx protein - African clawed frog

N:Alternate names: apical plasma membrane protein

C:Species: Xenopus laevis (African clawed frog)

C:Date: 30-Apr-1993 #sequence_revision 19-Oct-1995 #text_change 22-Jun-1999

C:Accession: A44361; S25517

R:Staub, O.; Verrey, F.; Kleyman, T.R.; Benos, D.J.; Rossier, B.C.; Kraehenbuhl, J.P.

J. Cell Biol. 119, 1497-1506, 1992

A:Title: Primary structure of an apical protein from Xenopus laevis that participates in

A:Reference number: A44361; MUID:93107151; PMID:11334959

A:Accession: A44361

A:Molecule type: mRNA
A:Residues: 1-1420 <STA>
A:Cross-References: EMBL:Z14997; NID:G64551; PID:CAA78178.1.1; PID:g64552
A:Experimental source: A6 cells
A>Note: sequence extracted from NCBI backbone (NCBIP:121141)
C:Complex: This protein is part of a large molecular complex.
C:Function:
A:Description: may be the amiloride-sensitive component of the amiloride-sensitive sodium C:superfamily: amiloride-sensitive sodium channel Apx protein
C:Keywords: glycoprotein; membrane-associated protein; sodium transport
F:119,462,481,503,660,664,988,1038,1211,1273/Binding site: carbohydrate (Asn) (covalent)

Query Match 3.9%; Score 161; DB 1; Length 1420;
Best Local Similarity 19.8%; Pred. No. 0.88;
Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

QY	42	EGINABQIVIKITDQGVYVTSHGDIY-----HYNGKVPY-----DAIISEE	82
Db	26	ERLSPVRSMTLLVDSAYSSFGSGSYVPEYQNSFOHGDCHYNDQQLSYMDSEYVRAIYNPS	85
QY	83	LLMKDPNYKLLKEDI VNEVGKGYIVKIDGK-----YVYLKDAAHADNVTK---	129
Db	86	LLDKDGVY-----NDIVSE-HGSSKVALSGRSSSLCSNTTTSVHRTSPAKLDNYVTNLD	140
QY	130	-----EETN-ROKQHSQHREGGTPND--GAVALARSQGR-----YTTDGGYIFNA	173
Db	141	EKNYIGDPLNMKHKQNRPNHKAYGLQRNSPTGINSLOEKENQLYNPSNFMFIKDNFYGRS	200
QY	174	SDIIEDTGDAYIVPHGDHYH---YIPKNELASELAAAEFLSGRNLNSR---TYR	225
Db	201	LDVLQADGB---IMTQDSTYONALYPQNPQDQVR---NTQYFGAARMSKEQKFNVDVQ	253
QY	226	RQNSDNTSRTNWPSYNSFGTTNTNTNSNTNSQASQSDNIDSLIKOLYKLPILSORHVE	285
Db	254	KSNEENTERDG--PYLTKDQGF-----VOGQYASDVRTSPKNI-----RRSLK	294
QY	286	SDGLVPDPAQIITSRTARG---VAVPHGDHYHP-----IPYSQMBELEERII---	331
Db	295	KSA-----SGKIVAHDSQSGWIMKPGKDTSPFNSEGTITDMYDNRQWDIRKSRLS	350
QY	332	--PLRYRNHNHWPDSPRPEQSPQPTPEPSGPQAPNLKIDSNSLSVLVRKVGEGYVF	389
Db	351	SQSLYYENEDV-----SGPPLKAMNSKNEVDQTLN-----	381
QY	390	BEKGISRYVFAKDLPSYTNKLSKQESVSHTLTAKENVAPRQDFYDKA-----Y	444
Db	382	-----FKDATVKSIPLLSQQLQOEKCKSHPL-----SDLNCEKITKASTPMY	425
QY	445	NLLTEAHKALP-----XNKGRNSDPQALDKLERLNDESTNKEKLVDDLAFAPIT	499
Db	426	HLAGRHSAFIAPVENTNPAAQOEKLEKSKTLERMNNIS-----VLQSEPRPDNH	476
QY	500	RLGPNISOIETDEVRITQAQADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHHWTKD	559
Db	477	KLPNKS-----ITQLADLHDSVEG-----GNSGNLNSABE	508
QY	560	SLS-----DKEVAAQAYTKEGI-----LPPSPDADVKANP-----	591
Db	509	SLANDYIEKLKVAQKVLRETSEFRKDLQMSLPCREKLPAPKPEPTIDHFRSYSSSSANEE	568
QY	592	-----TGDSAAAIY-----NRVKEKRIPLVRLPYMVEHTVEVRKNGNLIIPH	635
Db	569	SAYLQTKNSADSSYKKDKDTEKVAVTRIGGRKRI-----TKEQKLCYSEPEKLD	617
QY	636	H--YHNITKFAWPDHHTYKAPNGYITLEDLPATIKYYVEHPDPERPHSDNGHGNASEHVL	693
Db	618	HLGTQKSNFAWKEEPTFANRREMSDSDISANRIKYLE-SKERTNSS---SNLSKTELKQI	673
QY	694	DH-----SDDPNKNFKADEEPEETPA-----EPEVPQVETEKVAQLK---	732
Db	674	QHNALVQYMERKTIQRPNSN---POVQWERTSLGPNYNEWSIYSETSSSDASQKYLRR	730
QY	733	-----EAEVLLAKVT-----DSSLKANATETLAGRLNNLTLLQIMDNNNSIMAEAKLLAL	781

QY	459	GRNSDFQALDKLIERLN---	DEBSTEKELVDDL---	LAFIAPT---	THPERLGRPN	505
Db	2065	IREQDRKQVETVLROYEIPPSDLNQSNIHPDVKVAVIELARLWPLYFNQVVEVVEKRPD	2124	:	:	:
QY	506	SQIE---	YTEDSVRIA---	OLAKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWI	556	:
Db	2125	ESVSTTIFALSEHGIRLIVHTPHDLENPLKIQDFFPETIADVSLEANDILSVHVRH---	2180	:	:	:
QY	557	GKDSLSDEKVAQAQYTKK---	---	GILPPSPDADVKAN---	PTGDSAA	597
Db	2181	----	EDEENAYSAVRIKTNQAQIKYTLDRCLSGGVPKRKFVLARTGREIPPGVN---	2232	:	:
QY	598	AIYNRVKGEKRIPLVRLPYVVEHTVEYKNGNLIIIP	----	HKDHYHNIKFAMFDDDTYKA	652	:
Db	2233	WLYGKIEN-----	RFGLLAQYVDTGDGNVPPIRHETSEDRDERVRP--	FDDEVPFS	2283	:
QY	653	PNGYTTLEDLFATIKYYVEHDERPHSNDGNGNASEHVLGKKDHSEOP--	NKNFKADEEPVE	711	:	:
Db	2284	SERYTMD--FAT--KYFRKPKDKKQETWAMEDISQIV-----	RSEKPIISOSLIADLIGNEE	2337	:	:
QY	712	ETPAEPVEPVQVETEBKVAQLKEAB-----	VILAKVTDSSLKANATETLAGLRNNLTL	763	:	:
Db	2338	SKYAVETFHAIKMFGEDEPLKKSSEMTDVVFKVLLICHROPTLR--	DEVYCOLIKQTS	2394	:	:
QY	764	QIMDNNSIMABAEKILALLKSGNSPSSVS	791	:	:	:
Db	2395	NISOKPSNLSAFWRLLTIIATYFSSILT	2422	:	:	:

RESULT 19

T18444
hypothetical protein C0385C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18444
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18444
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1650 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
C;Genetics:
A;Map position: 3
A;Introns: 1597/3; 1625/3
A;Note: C0385C

Query Match	3.8%;	Score 158;	DB 2;	Length 1650;	
Best Local Similarity	17.7%;	Pred. No. 1.6;			
Matches	146;	Conservative 127;	Mismatches 311;	Indels 242;	Gaps 34;
Qy	64	DHYHYNGKVPYDAIISEELMKOPNYKLKDEDIVNEKGGWIVKDGKYYVYLKDAHA	123		
Db	11	DNVHYTKQV-----IRYDHNNIKANSHLKQSEY-----TTEN	47		
Qy	124	DNVR--TKEEINROKQHSQHREGGTPRNDGAVA-----LARSQGYT	164		
Db	48	DEKGRWITSNNNTNNIQTHA-----NVKDDKSVENKTTTLINKDINEITWISKQDNYT	102		
Qy	165	TDDGYIFNASDIIEDTDGAYIVFHGDHYHYIPKNELSASELAABAFIPLSGRLNSNSTY	224		
Db	103	KYNCLFNFKSDVMRD-----YVRYSKEDSSNNMGNNGNNMGNMGNMNNM	150		
Qy	225	RRQNSDNTSRTNWVSVPNGPTTNTSNNSNTNSOASQSDNDISLLAQLYKLPLSQRHV	284		
Db	151	NGNNNMN-----GNNMNGNNMNGNNMNDNNMNG-----NNI	185		
Qy	285	ESDGLVEDPAQITTSRT--ARGVAVPHGDHYHPIYSQMSLEERIARIPLRYRSNHWVP	342		
Db	186	MNGNMNGNNMNNRNNMNGNMNGNN-----NINRNNIMVR-----NNITNNHHI	234		
Qy	343	DSRPEQSPQPTPESPGPAPNL-----KIDNSLSVQ-----	378		

Db 235 H I S N K T S L I E N K K K E E N I P F S S A N L Y K E M V N V K G S S D T P L L Y K R K I D K D N L S K K K R F 294
QY 379 --L V R V K G E G Y F E E K G I S R-----Y Y F A K D L P S E T V N L E S K L S K Q 418
Db 295 N Y L F N K D G E H F V K E N V Q N N I I D D D D D D D N H H D N V V V Y D V K E M E N E N K N K G S V K E 354
QY 419 E S V S H T I T--A K E N V A P R D--Q E F Y D K A Y N L L T E A H K A L F N K N G S N S D F O A L D K L L E F 473
Db 355 D G L H N V L V E L R N K D N L V N D N I N K S F E K N N I L Y I K T S D S L E N Y N E R K I Y K E I N K-----410
QY 474 L N D E S T W K E K I V-----D L L A F L A P I T H P E R L G K P N S Q I E Y T B E V R I A Q L A D K Y T S D 528
Db 411 --R E Y S N K N E V V H F K N D D-----S S I K K N S S E C L D E Q K K T Y--K Y T I I E 454
QY 529 G Y I F D E H D I I S D E G A V Y T P H M G H S H W I G K D S L S D K E-----V A A Q Y T K E G I L P P 581
Db 455 Q K R Y N F N D--R D N N N A Y I-----K D D T H K E K G Y Y L N I M V O S E E Y--K Y G S N N K 500
QY 582 S P D A D V K A N P T G D-----S A A A L Y N V K G E K R I P L V R L P Y M V E H T V E V K N G--627
Db 501 M D E M E I N Q H T N D F N I N E N L N K I Y F D D Y E G D E K K K K L D-----D H I Y T Q Q E Y K N N I 556
QY 628 N L I I P H K D H Y N I K F A W F D D H T Y K A P N G Y T L E D I F A T I K Y V E H P D R P H S D G W G N A S E 687
Db 557 N D I L--K D H L N-----D K E T K K K N E I E E--E K K N I E E E K K N K I E-----I E E 601
QY 688 H V L G K K H S E D P N K F A D E P V E T P A E P V Q V E T E K-----V E A O L K 732
Db 602 E K N K I E T E E E K K K I E E E E K K K I E M E E E K N I D D E K K N Y A N D K I I S H I D N V N C I K 661
QY 733 E A E V L L A K V T D S S L K A N A T E T L--A G L R N N L T Q I M D N N S I M A E A E K 777
Db 662 -I D A L L D H I E E K K K T G H K E I N L Y K E I K E Y Q K M L N D E N S I M L E H E K 706
RESULT 20
S73361
dnaJ homolog protein C09_orf910 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 10-Dec-1999
C:Accession: S73361
R:Himmelfreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma
A:Reference number: S73327; PMID:97105885; PMID:8948633
A:Accession: S73361
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-910 <HIM>
A:Cross-references: EMBL:AF000004; GB:U00089; NID:g1673671; PIDN:AAB95683.1; P
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, Novemb
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma heat shock protein dnaJ homolog C09_orf910; dnaJ ami
P:7-71/Domain: dnaJ amino-terminal homology <DNJ>

RESULT 20

A:Strain: ATCC 29342
 A:Variety: ATCC 29342
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 10-Dec-1999
 C:Accession: S73361
 R:Hammelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma*
 A:Reference number: S73327; MUID:97105885; PMID:8948633
 A:Accession: S73361
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-910 <HIM>
 A:Cross-references: EMBL:AE000004; GB:U00089; NID:g1673671; PIDN:AAB95683, 1; P
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, Novemb
 C:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: Mycoplasma heat shock protein dnaJ homolog C09_orf910; dnaJ ami
 E:7-71/Domain: dnaJ amino-terminal homology <DNJ>

Query Match	3.8%;	Score 157;	DB 1;	Length 910;
Best Local Similarity	19.0%;	Pred. No. 0.77;		
Matches 161;	Conservative 127;	Mismatches 305;	Indels 254;	Gaps 444
QY	66	YHYNGKVPYDAIISEEL-----LMKDP-----NYKLKDEDI NEVKGGVIVKVD--GKTY 114		
Db	34	YHEDRNKAPDRAQIPAEINEANDVLSPNPKRANYKYGHGVDN-EGGFQAFQADVDSFF 92		
QY	115	VYLKDAAHADNVRTKEEINRQKQEHSSQHREGGTPRNDGVAVALARSGRVTYTTDDGYFNAS 174		
Db	93	BEIEKSGAFDNL--SESNTKKKTKTKKKGW-----FMGKSQEESTSDTTEY--A 140		
QY	175	DIIEDTGDAIYVPHGDHYHYIPKNELASELAAAEFLSGRGNLSNSTRYRRQNSDNTSR 234		
Db	141	DY--DAGLEDYFPQSDYDPDIPVDVARIEBVPQS--AYADDIPVDVADGMWEQNAEVNSA 197		

QY 235 TNWVSVNSPGTTNTNTNNTSNTNSOASGNDIDSL- - - - - 271
Db : : : : :
198 SEIIPDV- - - - - DAGLDEFNTISSAAPQASDWEAMIGNPEYCFDAAGWNWKGFDEA 251
QY 272 -KOLYLPLSORHVESDGVLFPDAQITSTARGAVPHGDHYHFIPYSOMSELEERIARI 330
Db : : : : :
252 GOWWLLEETSPSSVNDETTSDDAVTAAT- - - - -VEETDQ- - - - - 288
QY 331 IPLRYSNEHWPDSDPE- - - - - QPSQPPTP- - - - - EPSPOP- - - - - OPAPNLKI 369
Db : : : : :
289 - - - - - DSWTANSAEPVDVETVPQLQETEPEPIITLSSBPFEAPASVWJEPTPEIBE 341
QY 370 DNSLSIVSQLVRKVGSGYYFEEKGISRYVFADLPSETVKLNLESKLSKOESVSHLTAKK 429
Db : : : : :
342 TTSAVEMDASVK- - - - - ADVSDEADATNEPTEQDTISEPEQ- - - - - ETDAALALEEINH- - - - - TADL 394
QY 430 ENVAERDOFFDYKAYNLLTE- - - - - AHKALFNKGRNSDFQALDKLLRLNDESTNKKLVD 487
Db : : : : :
395 E- - - - - PAEV- - - - - SATNDLEQDVVEKVNFSEPESTVDTAATDPVVEQAVETSTNGPKFF-N 446
QY 488 LLAFL- - - - - APITHPERLGKPNISOI- - - - - EYTEDEVRIAQLA- - - - - 521
Db : : : : :
447 FSSFVLSDQNPNPQTTHHEEDAAAEPEFTVDSGESTAPEVTIAESTVELEETAENNP 506
QY 522 - - - - - DKYTSDGYIFDHDHIISDEGDAYVTPHMHSWIKDLSLS 562
Db : : : : :
507 ATFVEYLQPTKTTVVDVKLDPETKVAPTVSSENSVAPEPEFVAGPEQTFS-W-KPAIS 563
QY 563 DKEKY- - - - - AAQAVTKEKGILPPSPDA- - - - - DVKANPTGDSAAA- - - - - IY 600
Db : : : : :
564 ETEETPLTAVEPASSTQILIADVTSPVPTATAIPAPSINAVPTAPVAETFEAAVDFLK 623
QY 601 NRVKGEKRPLV-RUPYMVEHTVVEKGNGLIIIPHDXDHYNIKFAWFDDOHTYKA PNGTYLE 659
Db : : : : :
624 EAAKIEAQLPVTPVPEQIDGT- - - - - DPSLL- - - - - TWDRYLEKTRK-LFH 665
QY 660 DLFIATIKYVVEHPDERPHSNDGWGNASEHLVKKDHS- - - - - DPKNFKADEPEVETPAEP 717
Db : : : : :
666 KLFLT- - - - - EQLP- - - - - FIVKTDQFEIVDPN- - - - - LDEHNVLIVTE- 700
QY 718 EVPQVETKEVAQLKEAEVLLAKVTDSSLKANAVETLAGLRNNLTLOIMDNNSIMAEAEK 777
Db : : : : :
701 HVPQICP-INEQLKBIR-YTRKLVDPOTQVTTTE- - - - - SITLEVQSHKSQTEA- 748
QY 778 LLALLKG 784
Db : : : : :
749 -IAIFKG 754

RESULT 21
S55101
hypothetical protein YMR219w - yeast (*Saccharomyces cerevisiae*)
N;Alternate names: hypothetical protein YM8261.i3; hypothetical protein YM9959.01
C;Species: *Saccharomyces cerevisiae*
C;Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
C;Accession: S55101; S57587
R;Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A;Reference number: S55089
A;Accession: S55101
A;Molecule type: DNA
A;Residues: 1-711 <DED>
A;Cross-references: EMBL:Z49809; NID:g854459; PIDN:CAA89934.1; PID:g854471; MIPS:YMR219w
A;Experimental source: strain AB972
R;Skellton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A;Reference number: S57587
A;Accession: S57587
A;Molecule type: DNA
A;Residues: 608-1658 <SKE>
A;Cross-references: EMBL:Z49939; NID:g887599; PIDN:CAA90190.1; PID:g887600; MIPS:YMR219w
A;Experimental source: strain AB972

[illegible]

C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 06-Feb-1998
C;Accession: S61103
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; W
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of *Saccharomyces cerevisiae* chromosome XVI left arm.
A;Reference number: S59677
A;Accession: S61103
A;Molecule type: DNA
A;Residues: 1-2195 <ALL>
A;Cross-references: EMBL:U41849; NID:gl147608; PID:gl147609; MIPS:YPL085W
C;Genetics:
A;Gene: SGD:SEC16
A;Cross-references: SGD:S0006006; MIPS:YPL085W

A:	Map position: 161
C:	Keywords: transmembrane protein
E:	F;1198-1214/Domain: transmembrane #status predicted <TM1>
F:	F;1250-1266/Domain: transmembrane #status predicted <TM2>
Query Match	3.8%; Score 157; DB 2; Length 2195;
Best Local Similarity	19.7%; Pred. No. 2.7;
Matches 124; Conservative 109; Mismatches 241; Indels 156; Gaps 31;	
Qy	188 HGDHYHVPKNEIASELAAAEFLSGRNLNSRTYRRQNDSMTSTTNWVPSVSNPGTT 247 ::: ::::
Db	29 HSEPLEPELTINSS-----FNDSVNRTESDIASKSD-VPPVS-----S 68 ::: ::::
Qy	248 NNTNS-----NNSNTNSQASQSNDI-----DSLKLKLYKLPL 279 ::: ::::
Db	69 STNISPANETOLEPTDQLHHKKLLNDSDQHDITADSNLDPNSIVEHSDSVITQT-KPAM 127 ::: ::::
Qy	280 SQRHVESGLVPDPAQIITSR-----TARGVAIPGDDHYHFIPYQMSE----- 322 ::: ::::
Db	128 SOEVEET-----AAHLSSRNPLDVAGELHNNEHTQKTAVSAVEDSDSFNEEGNH 181 ::: ::::
Qy	323 --LEBRIARIPLRYSHHWPDSPREPQSPOTPERPGQPAPNLKIDNSSLSVLQ 380 ::: ::::
Db	162 SIITSSLNDATPSQY--NHFLPSD-----GNLLSPELSGDTPHNVPLGTKN----- 228 ::: ::::
Qy	381 RKVGEGYVFEEKGISRYVFAKD-LPSETVKONLESKLKSQESVSHTLTKAKENVA PROEF 439 ::: ::::
Db	229 -EINDVCNDKEIS--LNANNVLPDELSKEDERLKLETHVS--TEEKQDIA--DOET 281 ::: ::::
Qy	440 YDKAYNLLTEAHKALFYXNGRNSD--FOALDKLLERINDESTNKLEKLVDDLAF LAPITH 497 ::: ::::
Db	282 AENLFSTSSTEPSENKIRNSGGDTSLMFG-----DDESQKVPEWEDVKKOF----H 328 ::: ::::
Qy	498 PERLGKPNSEQIEYTEDEVK-----LAOLADKYTTS DGVIPE---HDIISDEGDAYVTP 548 ::: ::::
Db	329 NENTNNTQESAPNTDDRDKYGEGNEALKXSCECTAADERSYSEESEDIFGHCDKQVVE- 387 ::: ::::
Qy	549 HMHGHSHWIGKDSLSDKEKVAQAAYTKEGILPPSPADV--KANPTGSAAIYNRVKGE 606 ::: ::::
Db	388 --GQNDFTCKNIENESQKLWGEGNHK---LPLSABADIIEPKDIIQDAEFLTQS SGD 441 ::: ::::
Qy	607 KRPIPLVRPLPYMVHEHTVEVNGNLIIPH KDHYNI-----KPAM-----FDDHYTKAPN 654 ::: ::::
Db	442 -----LGEVLFWESTDNRADVTYSQSKBEDLFAASGNDEKLPWEVS DGVSSGKTYEN 494 ::: ::::
Qy	655 G-YTLEDLFPATIKY-YVEHPDERPHENDCGWNASEHVLCGKHSDSPDNKNFKAD EEPVEE 712 ::: ::::
Db	495 SMOQTSTEKTAEQFSFLENDLDLDDDDDFLASSE-----EEDTVPTNDTNTLT SKPVEE 550 ::: ::::
Qy	713 TPAEPVPPQVETEKVEAQLKEAEVLAKVT 742
Db	551 KKASRYKPIIE---BAGMRQEOVHFTNITT 577

RESULT 23
T48429
hypothetical protein F8F6_230 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #
C/Accession: T48429

R;Bavan, M.; Peters, S.A.; van Staveren, M.; Dickse, W.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24488
A;Accession: T48429
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1495 <BEV>
A;Cross-references: EMBL:AL162873
A;Experimental source: cultivar Columbia; BAC clone F8F6
C;Genetics:
A;Map position: 5
A;Introns: 100/3
A;Note: F8F6.230

Query Match	3.8%	Score 156.5	DB 2	Length 1495
Best Local Similarity	18.6%	Pred. No. 1.7	Indels 219	Gaps 35
Matches 146	Conservative 124	Mismatches 297		
QY	173	ASDIIDTGDAYIVPHGDHYHYIPKNELSASELAA--AEAFLSGRGNLSNRTYR-----	225	
DB	716	SDEALESTADASLC-----NHLAVEEVDGLGCSFIEEEKKGESEKQNLSTWRNLQK	770	
QY	226	-----RQNSD-NTSRT---NWVSVSNPGTNTNTNSNS-----	NTN 258	
DB	771	HMWRDSEGNRNETQECHKW---SYGTDQMTGDDANAATAVAKIQIAFATILSEIPDSSS	828	
QY	259	SOASQNDIISLKKO-----LYKLPLSQRHVESDGLVDFPAQITS-RTARGV	304	
DB	829	DEESVSSNSLKEEKEHQGETKSWNSLRKVLKFFVKS---LEKVQVNPVKRKNRL	884	
QY	305	AVPHGDHYH--FIPYSQME-----LEBRIARIPLRYR-----	336	
DB	885	PVESAFEAENVFLRHRSIMEGRTTEGEMMLDYALROAISRLAPQIKKQVLLVQAFDIV	944	
QY	337	SNHWVDSRPEQSPQPTPEPSG-POPAPNLKIDSNSSLSVSLVKVKGEGVVFEEKGI	394	
DB	945	LDGHDTPKQTSNPTPRNNDTEKGPKEVVEGCEVKKD---EQIKNVFARQVHQKDL	1000	
QY	395	SRYVFAKDLPSGT-----VKNLESKLSKQES-----VSHTLTAKK-----	ENVA 433	
DB	1001	KGEEVHNTPKESRNLPFRNFQRIVVEKGDSRMWKLIVKHMVTEKEGIDGSANARSA	1060	
QY	434	PRDQFYDKAVNLLTEAHKALPXNKGNSDFQALDKLLERLNDSTNKEKLVDDLLAFLA	493	
DB	1061	SVESYDDEAGGLQIDARRSGTVTLVR---EALEKILSEIPDNSSDDQMSD-----	1110	
QY	494	PTTHERLKGKNSQIETVEDVRIAQLADKYTTS--DGY-----IFDEHDIISDEG---	543	
DB	1111	ITTEQELFERNQSVSEKSESVSSATPKPKETKRVKGNVNVKVKILLKRFVSDLSGLTR	1169	
QY	544	AYVTPHM-----GHSHWICKDSLSDEKVKVAAQAYTKKGL	579	
DB	1170	LSPKTPRVLPWEPDETEKIRLHQEIGCKRNSBEMMLDYALROAISTLAPSORRKVSLL	1299	
QY	580	PPSPDA-----DVKANPTGDSAAIIVRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHK	635	
DB	1230	AQAFDTISLQMGSGSTPCGSAASRN--ISROSSISSMAHYENANAEIIRGKL-----	1282	
QY	636	HYHNTKFAWFDHTYKAPNGVYTLDFATIKY--YVEHPDRPHSNDGWNASHEVLGKK	693	
DB	1283	--RNLQ-----EDLESAKLDGVSKDLEEKQCGSSLW-----RILCKQ	1318	
QY	694	DHSEDQNNFKKADPEPVETPAEPVQVETKEVAQLKEAVLLAKYTDS-SLKANATE	752	
DB	1319	MEDNEKNQPLPETKEEELKEDTSDVGEGKMELYOTEAVELLIGEVIDGISLEESQDQ	1378	
QY	753	TLAG--LRNLLTLQI---MDNNSIMAEA-----EKLALLKGSNP-----	SSV 790	
DB	1379	NLNNETROKSETLQVSKVIRIDRWSNLKRAILLRRFVKALENVRKFNPREPFLPNPEV	1438	
QY	791	SKEIN	796	
DB	1439	EAEKN	1444	

us-09-765-271-56.rpr

Fri Oct 1 15:35:55 2004

Qy	578	ILPSPDADVKANFTGDSABAIIYNRVKCEKEIPIVRLPYMVEHTVEKNGNLIIPHKOHY	637
Db	940	VVEESPQYMHANYNNDIVHAV-----LYEPYNSNNVISYTN-----NKCHK	982
Qy	638	HNIKFAWF---DDHTYKAPNGYTLDFATIKYYVEHPDERPHSNDGWNASEBVLGK-	692
Db	983	NNFDINYLOHRNDNNSNKONISLTNDYTFNSDYIANMMDHDTYRKQIIKNIPAHQISKL	1042
Qy	693	KDHSSEDPNKNFKADEEPVE	711
Db	1043	KD-----GKNFKAYNESIQ	1056

Search completed: October 1, 2004, 07:17:16
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 03:36:05 ; Search time 19 Seconds
(without alignments)
2181.465 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTKENRVSY.....KLLALLKGNPSSVSKEKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	4.1	1233	1 YF16 YEAST	P43597 saccharomyc
2	161	3.9	1420	1 AXP_XENLA	Q01613 xenopus lae
3	158	3.8	1118	1 UBPH HUMAN	P40818 homo sapien
4	157	3.8	910	1 DNJM MYCPN	P75354 mycoplasma
5	157	3.8	1658	1 YM67 YEAST	Q03661 saccharomyc
6	157	3.8	2195	1 SC16 YEAST	P48415 saccharomyc
7	156.5	3.8	1185	1 MAPX DROME	P23226 drosophila
8	154.5	3.7	3924	1 ANK2 HUMAN	Q01484 homo sapien
9	154	3.7	1453	1 Y373 BOVIN	Q9tu23 bos taurus
10	153.5	3.7	1539	1 Y373 HUMAN	O15078 homo sapien
11	153.5	3.7	2867	1 RBP2 PLAVB	Q00799 plasmodium
12	150.5	3.6	1176	1 YOH8 YEAST	Q08236 saccharomyc
13	149.5	3.6	1175	1 PTNL RAT	Q62728 rattus norv
14	149	3.6	2869	1 REP1 PLAVB	Q00798 plasmodium
15	145.5	3.5	1085	1 CUT7 SCHPO	P24339 schizosacch
16	145.5	3.5	1240	1 YNJ1 YEAST	P53935 saccharomyc
17	144	3.5	1337	1 JIP3 MOUSE	Q9esn9 m c-jun-ami
18	143.5	3.4	900	1 SVA MYCGE	P47534 mycoplasma
19	142.5	3.4	875	1 Z1P1 YEAST	P31111 saccharomyc
20	142.5	3.4	1066	1 STS5 SCHPO	O4454 schizosacch
21	142	3.4	850	1 D7 DICDI	P54682 dictyosteli
22	142	3.4	1744	1 TANX XENLA	Q01550 xenopus lae
23	141.5	3.4	490	1 MOT3 YEAST	P54785 saccharomyc
24	141.5	3.4	1928	1 MYSL YEAST	P08964 saccharomyc
25	141.5	3.4	2314	1 PTPZ HUMAN	P23471 homo sapien
26	141.5	3.4	2748	1 NUM1 YEAST	Q00402 saccharomyc
27	141	3.4	914	1 IP42 YEAST	P39936 saccharomyc
28	141	3.4	1332	1 SP7 YEAST	P35177 saccharomyc
29	140.5	3.4	516	1 P54 ENTFC	P13692 enterococcu
30	139.5	3.3	1609	1 LMGI HUMAN	P11047 homo sapien
31	139	3.3	1957	1 SPOF SCHPO	Q10411 schizosacch
32	138.5	3.3	1703	1 SNF1 YEAST	P22082 saccharomyc
33	138	3.3	1639	1 MSPI PLAFW	P04933 plasmodium

34	137.5	3.3	1790	1 USO1 YEAST	P25386 saccharomyc
35	137.5	3.3	2175	1 HMCU DROME	P10180 drosophila
36	137.5	3.3	2459	1 MAPB RAT	P15205 rattus norv
37	137	3.3	1176	1 PTNL MOUSE	Q62136 mus musculu
38	136.5	3.3	679	1 YI83 YEAST	P40563 saccharomyc
39	136.5	3.3	1183	1 CNA STAAU	Q53654 staphylococ
40	136	3.3	1630	1 MSPI PLAFK	P04932 plasmodium
41	136	3.3	1836	1 MSPI YEAST	P25558 saccharomyc
42	135.5	3.3	1053	1 SLPM BACBR	P06546 bacillus br
43	135.5	3.3	1976	1 MYHA BOVIN	Q27991 bos taurus
44	135.5	3.3	3358	1 PGCW MOUSE	Q62059 mus musculu
45	135	3.2	1532	1 IGA NEIGO	P09790 neisseria g
46	135	3.2	3664	1 MINT HUMAN	Q06458 homo sapien
47	134.5	3.2	960	1 YMX6 YEAST	Q04279 saccharomyc
48	134.5	3.2	1124	1 TCF8 HUMAN	P37275 homo sapien
49	134	3.2	1163	1 RTN4 RAT	Q9Jk11 rattus norv
50	134	3.2	1188	1 OSH1 YEAST	P35945 saccharomyc
51	134	3.2	6632	1 UN89 CAEEL	Q01761 caenorhabdi
52	133.5	3.2	1607	1 LMGI MOUSE	P02468 mus musculu
53	133.5	3.2	1664	1 INT1 CANAL	P53705 candida alb
54	133	3.2	1189	1 YOH6 YEAST	P47035 saccharomyc
55	132.5	3.2	1875	1 MLF1 YEAST	Q02455 saccharomyc
56	132	3.2	1164	1 TSCI HUMAN	Q92574 homo sapien
57	132	3.2	2431	1 POLN SFV	P08411 senlixi for
58	131.5	3.2	764	1 MSP1 YEAST	P54199 saccharomyc
59	131.5	3.2	1046	1 YU14 YEAST	P47029 saccharomyc
60	131.5	3.2	1379	1 YFF9 SCHPO	O14066 schizosacch
61	131	3.1	1014	1 HEX2 YEAST	Q00816 saccharomyc
62	131	3.1	1093	1 TWF1 HUMAN	P82094 homo sapien
63	131	3.1	1358	1 SIR4 YEAST	P11978 saccharomyc
64	131	3.1	1612	1 DNM1 PARLI	Q27746 paracentrot
65	130.5	3.1	704	1 MSN2 YEAST	P33748 saccharomyc
66	130.5	3.1	1547	1 TOP2 BOMMO	O16140 bombyx mori
67	130.5	3.1	1849	1 IGA4 HAEIN	P45386 haemophilus
68	130.5	3.1	2805	1 MAPA HUMAN	P78559 homo sapien
69	130	3.1	969	1 SABC STRSL	Q55242 streptococc
70	130	3.1	1283	1 OSH2 YEAST	O12451 saccharomyc
71	130	3.1	1829	1 RM1 MOUSE	Q70472 mus musculu
72	130	3.1	6669	1 NEBU MOUSE	P20929 homo sapien
73	129.5	3.1	1001	1 REPR MOUSE	Q9r0x5 mus musculu
74	129	3.1	646	1 SG1 BOVIN	P23389 bos taurus
75	129	3.1	1026	1 STAÜ DROME	P25159 drosophila
76	129	3.1	1031	1 YDG9 SCHPO	Q10496 schizosacch
77	129	3.1	1570	1 P3K1 DICDI	P54673 dictyosteli
78	129	3.1	2278	1 FABI YEAST	P34756 saccharomyc
79	129	3.1	3038	1 TRIO HUMAN	O75962 homo sapien
80	128.5	3.1	1137	1 MSB1 YEAST	P21339 saccharomyc
81	128	3.1	742	1 CD44 HUMAN	P16070 h cd44 anti
82	128	3.1	817	1 YG4A YEAST	P46949 saccharomyc
83	128	3.1	1039	1 MSLL DROME	P50535 drosophila
84	128	3.1	1616	1 P200 MYCGE	Q49429 mycoplasma
85	127.5	3.1	782	1 Y044 UREPA	Q9pral ureaplasma
86	127.5	3.1	954	1 BIR1 YEAST	P47134 saccharomyc
87	127.5	3.1	1334	1 JIP3 HUMAN	Q9upt6 homo sapien
88	127.5	3.1	1466	1 SPA2 YEAST	P23201 saccharomyc
89	127.5	3.1	1597	1 GTFI STRDO	P11001 streptococc
90	127.5	3.1	1608	1 HLYA SERMA	P15320 serratia ma

ALIGNMENTS

RESULT 1	YF16 YEAST	STANDARD;	PRT; 1233 AA.
ID	YF16 YEAST	AC	P43597;
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Hypothetical 137.7 kDa protein in UGS1-FABI intergenic region.		
GN	YFR016C.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		

Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 (1)
 SEQUENCE FROM N.A.
 STRAIN=S288c / AB972;
 MEDLINE=95400292; PubMed=7670463;
 Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 Yamazaki M., Tashiro H., Eki T.;
 "Analysis of the nucleotide sequence of chromosome VI from
 Saccharomyces cerevisiae";
 Nat. Genet. 10:261-268 (1995).
 -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL; D50617; BAA09255.1; --
 PIR; S56271; S56271.
 Germline; 140170; --
 SGD; S0001912; YFR016C.
 Hypothetical protein.
 SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;
 Query Match 4.1%; Score 171; DB 1; Length 1233;
 Best Local Similarity 19.8%; Pred. No. 0.15;
 Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;
 QY 12 VKNRVSVDGKQATQKNTLNPDEVSKEGNAEQIVI-----KITD-QGVVTS 61
 Db 151 IKETSTNNVAEG-----TENVP--IKSTGIEVGNPSITRRKKNKKKTTNRRGRNS 202
 QY 62 HGHYHYNGKVPYDAII--SELLMKDPNKLKDEI-----VNEVGGYVTKVD 110
 Db 203 NPADITDLSKQSTLSDSLVIGIEYLQEDGS---KNEDIKVINVDPEVNVKMDIRTRNE 259
 QY 111 GKVVYLKAAHADNV-----RTKEENRQKQHSQHREGG--TPRNDGVALARSO--- 160
 Db 260 SSKTFDIDVKNKDNDVETSKSENNINEEKAHTLPREENEILNVNEGNAASFQKQLEP 319
 QY 161 -GRYTTDDGYFNASDIIETGDYIVPHGDHYHYPNLSASELAAAFISGRNLS 219
 Db 320 HGLEAGDE---NGQASTKDVESRLTKNGFNFKENESKHLKAGEKQOTE---SDRDGIS 372
 QY 220 NSRTYRQNSDNTSRTNWPSVSNP-----GTNTNTSNNSTNSQASQNDI----- 267
 Db 373 PSVLAKNQKTEIGKEDHVFQKDEKCKRKLSDVNHNNMHNFNVAAGSDSIIPPETE 432
 QY 268 ----DILLKQYLKPLSQRV-----ESDGLVFPDPAQ---ITSR 299
 Db 433 REYVDETMGTPTKISDNKMLQHTNDISVEVEKEEEEEENSTFSKVKKENVTGE 492
 QY 300 -----TARGVAVPHGDHYHYPISQMSLEERTARILPIYRSN----- 338
 Db 493 QEAVRNVNVSSTESTSKGEIMGGD-----EKQSEAGEK-SSIIIEGSAKIS 544
 QY 339 --HWVPSRPPQSPQPTPEPSPGPQAPNLKIDNSLSVLYRVKGVGVFPEKGTSR 396
 Db 545 KDNVLDEEAPTOENKPTVEVGEIDIPDPRDD-----VEIIVAEKNIIPDELVAK 599
 QY 397 -----YVPAKDLPSVETKNSKLSKQSVSHITAKENVAPR-----DOEFYDK 442
 Db 600 EDQGEQVKLDEPVKMKDDKIAMEGABSISEDKKKQEGTAELSNEKAKKEVDETARES 659
 QY 443 AYNLLTEAHKALFNK-----GRNSDFQALDKLLELNDSTNKELKVLDDLAFLAPI 495
 Db 660 AEGVEVEKSKTPSPKPVKVRCTSGRPDLQINERPDVRLKEDVRVPDVEDVPEIATTIEN 719

496 TPERLKGPNQIEYTEDVIRIAQADKYTTSDGYVFEHD-----IISDEGDVAVTTPM 550
 Db 720 SEEDPKSQRVQISTEQAEITQKMDGVGTSITS---FKEEKPKRFEITQEGDKITGDT 776
 QY 551 GHSWIGKDSLSDKEKVAQAAYTKKGILPPS-----PDADVKANPTGSAALVNR 602
 Db 777 NHEHGEATEAASNSK-ASDVGTAEKYTFPSSSVKDKTEEDAEVE-----NSEKTEFIK 830
 QY 603 VKGKRIPIVRLPYVVEHTVEKNGNLI-----PHKDHVHNKIFAWFDDHTY 850
 Db 831 VKAE--LENLDAPKAEVTAELNKENEDVEVDTEEDAEVENSEKTEFIKVK---AELGNL 885
 QY 651 KAPN--GYTLE-----DLFATIKYVVE---HPDERPHNDGNGNASEHVLGKD-- 694
 Db 886 DAPKAEVTAELNKENEDVEVAATSKEDIETKCEPAETPIED---GTCEAEVSKKDAE 942
 QY 695 --HSEDPNKFKADPEVETAEAEVPOVE-----TEKVE-AQLKEAEVLLAKVTDSSL 746
 Db 943 AVTKEDENMENSKIABALKDVTGDQDQIDINISDEFQRTVELPELEKQDIKNGKEDKEL 1002
 QY 747 KANATETLAGL-----RNNTLIQIMDNNSIMAEKLLALLKGSNPSVSKKIN 796
 Db 1003 EVEETEKETSILPDVVEENIT---BEKNIKQEEEB-VSQLDNFNETESISKEAPN 1053

RESULT 2
 APX_XENLA STANDARD; PRT; 1420 AA.
 ID APX_XENLA
 AC Q01613;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Apical protein (APX).
 GN APX.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=93107151; PubMed=1334959;
 RA Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,
 Kraehenbuhl J.-P.;
 RT "Primary structure of an apical protein from Xenopus laevis that
 participates in amiloride-sensitive sodium channel activity.";
 RL J. Cell Biol. 119:1497-1506 (1992).
 CC -!- FUNCTION: Is part of a multimeric complex which is involved in
 amiloride-sensitive sodium channel activity.
 CC -!- SUBCELLULAR LOCATION: Membrane-associated.
 CC -!- TISSUE SPECIFICITY: Kidney, proximal intestine, oocytes, and
 to a lesser extent in the distal intestine, stomach and eye.
 CC -!- SIMILARITY: SOME, TO HUMAN APXL.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL; Z14997; CAA78718.1; --
 PIR; A44361; A44361.
 Membrane; Sodium transport; Transport.
 DOMAIN 111 114 POLY-SER.
 FT DOMAIN 559 564 POLY-SER.
 FT DOMAIN 1048 1051 POLY-THR.
 SQ SEQUENCE 1420 AA; 159467 MW; 0DD8B5C11413FFBC CRC64;
 Query Match 3.9%; Score 161; DB 1; Length 1420;
 Best Local Similarity 19.8%; Pred. No. 0.62;

Matches	180;	Conservative	111;	Mismatches	318;	Indels	302;	Gaps	44;
QY	42	EGINAEQIVIKITDQGYVTSHGDHY	---	HYNGKVVPY	----	DALISEE	82		
Db	26	ERISFVRSMTTLVD	SAYSSFGSSVPEYQNSFQHDGCHYDEQLSYMDSEYVPAIYNPS	85					
QY	83	LLMKDPNYKLDEIDVNEKGYVIKDGK	----	YVVLKDAAHADNVRTK	----	129			
Db	86	LLDKDGVY	----	NDIVSE-HGSSKVALSGRSSSLCSDNTTTSVHRTSPAKLDNYVTTLDS	140				
QY	130	-----EEIN-RQKEHSQHREGGTPRND	--GAVALARSQR	----	YTTDDGYIFNA	173			
Db	141	EKNYIGDPIINMKHKQNRPNHKAYGLQRNSPTGINSLOEKENOLYNPSNFWEIKDTPGRS	200						
QY	174	SDIILEDGDAIVPHGDHYH	----	YIPNELSASELAAAEAFISGRGNLSNR	----	TYR	225		
Db	201	LDVIAQAGD	----	INTQDSYTONALLYFPQNPQDYR	----	NTQYPGANRMSKEQFKVNDVQ	253		
QY	226	QNSDNTSRTNWVPVSNPGTNTNTNNSNTNSQASQSDIDSLKLQYKLPLSQRHVE	285						
Db	254	KSNEENTERDG	-PYLTQDGF	----	VQGYASDVRTISFKN	----	RRSLK	294	
QY	286	SDGLVDPDAQITSTARG	----	VAVPHGDHYHF	----	IPYSQMSSELEERARI	331		
Db	295	KSA	----	SGKIVAHDSQGSQWIMKPGKDTPSFNSEGTITDMDYDNREQWDIRKSRLSTRA	350				
QY	332	--PLRYBSNHWPDSPRPEQSPQPTPEPSQPQAPNKLKDSNLSVQLRVKVGEGYVF	389						
Db	351	SQSIIYESNEDV	----	SGPLKAMNKSNEVDQTL	----	381			
QY	390	EEKGISYVFAKOLPSETVKNLSKLSQBSVSHTLTAKKENVAPRDQDFYDKA	444						
Db	382	-----FQKDATVKSIPLLSQOLQEQEKKSHPL	----	SDLNCEKIKTASPTMPLY	425				
QY	445	NLLTEAHKALF	----	XNKGRNSDFQALDKLLERLADNESTNKEKLVDDLLAFIAPITHPE	499				
Db	426	HLAGRHSAFIAPVHNTNPAQQBKLEKSTLERMNNIS	----	VLQSEPRPDNH	476				
QY	500	RLGKPNISQIEYTEDEVRIAQALADKYTTSDGYIFDEHDIISDEGDAVVTTHMGHSHWIKD	559						
Db	477	KLPNKS	----	LTQLADLHDSVEG	----	GNGLNLSAAEE	508		
QY	560	SLS	----	DREKVAQAQYTKEGI	----	LPPSPDADVKAMP	591		
Db	509	SLMNDYIELKVAQKVLRETSFKRDLQMSLPCRFKLNPPKRPRTIDHFRSYSSSSANEE	568						
QY	592	-----TGDGSAAY	----	NRVKGEKRIPLVRLPVYVHTVEVNGNLIIIPHK	D 635				
Db	569	SAVLQTKNSADSSYKXDDTEKVAVTIGRKRI	----	TKEQKLCYSPEKLD	617				
QY	636	H--YHNKIFAWPDHHTYKAPNGYTTLEDLPATIKYVVEHDPDRPHSDNGWGNASEHVLGKK	693						
Db	618	HLGIQKSNFAWKEEPTFANRRMSSDSISANRIKYLE-SKERTNSS	----	SNLSKTLELQI	673				
QY	694	DH	----	SEDPNKNFRADEEPPVETA	----	EPEVPQVETEKVEAQIK	732		
Db	674	QHNAIVQYMERKTNQRPNSN	----	PQVMERTSLGUPNWNESIYSETSSSDASQYLER	730				
QY	733	-----EAEVLLAKVT	----	DSSLKANATETLAGLRNNLTQIMDNNSIMAAEKLAL	781				
Db	731	RSAGASSSYDAVTWNDRFGKTSPLGRSAAEKTAGVQRK	----	TFSDQRTLDGSOEH	783				
QY	782	LKGSNPSSVSK	792						
Db	784	LEGSSPSLSQK	794						

RESULT 3

UBP8_HUMAN

ID

TUBP8_HUMAN

STANDARD;

PRT; 1118 AA.

AC

P408T8;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

Db 174 KCETKEGAIITAKELYTMTDKNISLIIMDARMQDYQDSCL-----HSLSVPEEAI SPG 229
Qy 204 ELAA-AEAFSLGRGNLSRTYRRQNSDRTGRNWPVSVN--POTNTNTNSNNTSQQ 260
Db 230 VTASWIEAHL--DDSDTKWKRGNVEYVLLDFWSSAKDLIGTTLRSKL-----278
Qy 261 ASQNDIISLLKQLKPLSQHVESDG-----LVDPDAQITSTRARGVAVPHGDHYHFI 315
Db 279 -----DALFWESKTVLRNEPLVLEGGYENWLLCYQYTTNAK--VTPP-----320
Qy 316 PYSQMSSELEERTARIPLRYRNHWHVDSRPPSPQPTPEPSQPQAPNLKIDNSSL 375
Db 321 PRQ-----NEEVSISLDFTPS--LEESIPSKPAQ-----TPPASIEVDENIEL 364
Qy 376 VSQLVRKV-----EGYVFEKGISRYVFAKPLPSETVNL-----ES 413
Db 365 ISQNERMGLNISTDVEPVAASKSDVSPIL-----QPVPSIKNVQIDRTKPAVKLPPE 420
Qy 414 KLSQESVSHLTAKENAVPQDEFYDKAYN--LLTEAHK-----ALEXNKGRNSD 463
Db 421 HRIKSESTNHEQSQSQSGKVPIDRSTKPVVFSPTMLTDEEKARIIHAETALLMEKNKQB- 479
Qy 464 FOALDKLLRLNDESTNKEKLVNDLLAFLAPITHPERLGKPNISQI---BYTEDEVRIAQ 520
Db 480 -----KELRERQEE--QKEKLKE-----EQEQKAKKKQEAENEITEKQKAKEE 524
Qy 521 ADKYTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKLSLSDKEKVAQAQYKEKGIILP 580
Db 525 MEKKESEQAKKEDKE--TSAKRKEIT-----GVKROSKSHETSADKKSVEDRGKRC 575
Qy 581 PSPDADVKA-----NPTGDSAAIYNRVKGKRIPLVPLPYMVEHTVEVKNGLN-----629
Db 576 PTEIOKSGDGVHSTSVTSDSGSGKPFKIGQPESGIILRTGTFREDTDTTERNAQREP 635
Qy 630 -----IIPKDHVHNIKFAWDDHTYKAPNGYT--LEDLPATIKY-----VEH 671
Db 636 LTFARSEEMGRIVP-----GLPSGAKFLDPITGTFRYHSPNTVH 677
Qy 672 -----PDERHNSDNGWNASHEVLKGDHSDPNKNPKADEP-----VBETPAEPE 718
Db 678 MYPEMAPSAPPSTPTTHKAKTQIPABRD--REPSKLKESYSSPDITQAIQEEKKRPT 735
Qy 719 V-PQVETEKVEAQKAEVILAKVDTSSLK-----ANATETLAGLNR-----NLTLQ 764
Db 736 VTPVNRNKPCTYCPKAEI--SRLSAQIRNLNLPVFGSGFALTLGLNLGNTCYMNSILQ 793
Qy 765 IMDNNSIMAE 774
Db 794 CLCNAPHAD 803

RESULT 4

DNJM MYCPN STANDARD; PRT; 910 AA.
ID DNJM MYCPN AC P75354;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ-like protein MG200 homolog.
GN MPN119 OR MP035.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- SIMILARITY: Contains 1 J domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000004; AAB95683.1; --
CC PIR; S73361; S73361.
CC HSP; P25685; 1HDJ.
CC InterPro; IPR001623; DnaJ_N.
CC InterPro; IPR008971; HSP40 DnaJ_pap.
CC InterPro; IPR003095; Hsp_DnaJ.
CC Pfam; PF00226; DnaJ_1.
CC PRINTS; PRO0625; DnaJPROTEIN.
CC SMART; SM00271; DnaJ_1.
CC PROSITE; PS00636; DnaJ_1; 1.
CC PROSITE; PS00767; DnaJ_2; 1.
KW Hypothetical protein; Chapterone; Complete proteome.
FT DOMAIN 4 73 J-DOMAIN.
SQ SEQUENCE 910 AA; 100190 MW; 125D0E37D2D221A7 CRC64;

Query Match 3.8%; Score 157; DB 1; Length 910;

Best Local Similarity 19.0%; Pred. No. 0.55;

Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;

Qy 66 YHYNGKVPYDALISEEL-----LMKDP-----NYLKDSDIVNEVKGVIKVD--GKYV 114
Db 34 YHPDRNKAPDAQIIPAEINEANDVLSNPKKRYNYKXGHDGVDN--EGGFAPQADVDFSPF 92
Qy 115 VYLKDAHADNVRTKEEINRQKESQHREGGTPRNDGAVALARSGRYTDDGYTFNAS 174
Db 93 ERIEKSAGFNL--SESNTKKKTKTKKGW-----FWGSKQESTSDTTEY--A 140
Qy 175 DIIEDTGDAYVPHGDHYHVPKVELSASELAFAFLSGRGNLSNRTYRRQNSDTSR 234
Db 141 DV--DAGLEDYPPQSDYPPDDIPVDARIIEVDQS--AYADDIPVDAGMDQWQAEVANS 197
Qy 235 TNWVPSVSPGTNTNTNSNNTNSQASNDISLL-----271
Db 198 SEIIPDV-----DAGLADEFNTSSAPQASDWEAMIGNPEYGYFDAAGWNKGFDEA 251
Qy 272 -KQLYKPLPSQRHVESDGLVDPQAITSRTARGVAVPHGDHYHFI PYSQMSSELEERIAFI 330
Db 252 GQWVLEETEPSSVSNDEITTDSDAVTAAT-----VEETDQ-----288
Qy 331 IPLRYRNHWPDSRPE-----QPSQPTP-----EPSPGP-----OPAPNLKI 369
Db 289 -----DSWTANSAPPEVDVETPVELQPEPEPIITLSSPEVAPASVVIETPIEE 341
Qy 370 DSNSSLVSQLVKRGYGVFEKGISRYVFAKPLPSETVKNLSKLSKQSVSHLTAKK 429
Db 342 TTSAVEMDASVK-----ADVSDEADATNEPTEQDTISEPEQ--ETDAAALEINH-TADL 394
Qy 430 ENVAPRDQEFYDKAYNLLTE--AHKALFXNKGNSDFQALDKLLERLNDSTKEKLVDD 487
Db 395 E---PAEV---SATNDLEQDVVEKVFSEPESTVDTAATDPVVEQATETSTNGKFPF--N 446
Qy 488 LLAPL-----APITHPERLGKPNISQI-----EYTEDDEVRIAQ 521
Db 447 FSSFVLSDQNPNPQTPTTHHEEADAAPETVDETSGESTAPEVITAEVLETAEEINNP 506
Qy 522 -----DKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKLSLS 562
Db 507 ATFVEEYLOPTKTTVVDKLEDEPTVAKPTVSDSENSVAPEPEFVAGPEQTES--W--KPAIS 563
Qy 563 DKEKV-----AAQAYTKEGILPPSPDA-----DYKANETGDSAAA-----IY 600
Db 564 ETEBIPLTAVEPASETQTLIAEDVTSFVPTTATAPAPSINAVPTAPVAETFEAAVDLCK 623
Qy 601 NRKVGKRIPLV-RLPYMVEHTVEVKNGLIIPKDHVHNIKFAWDDHTYKAPNGYTL 659

```

Db 624 EAAKIEAQLPVTPTVEQIDGT-----DPSLL-----TQWDEYLEKTRK--LFH 665
QY 660 DLFAATIKYVVEHPDERPHSDGNGWGNASEHVLGKKHSE--DNKNPKADEEPEVETPAPP 717
Db 666 KLFLF-----EQLP-----FIVKTDQFEIVDEN-----LDEHNVNLIYTE- 700
QY 718 EYPQVETKVAQKAEAEVLLAKVDSSLLKANATETLAGLRNLTLOIMDNNSIMAAEAK 777
Db 701 HVPQICF--LNEQLKEIR-YTRKLVDPQTVTTT-----SITLEVQLSHKSQTEA-- 748
QY 778 LLALLKG 784
Db 749 -IAIFKG 754

RESULT 5
YM67_YEAST
ID YM67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUAL-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin M., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome
XIII."
RL Nature 387:90-93 (1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z49809; CAA89934.1; -
DR EMBL; Z49939; CAA90190.1; -
DR PIR; S55101; S55101.
DR GerMOnline; 142894; -
DR SGD; S0004832; ESC1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.8%; Score 157; DB 1; Length 1658;
Best Local Similarity 19.7%; Pred. No. 1.2;
Matches 176; Conservative 119; Mismatches 338; Indels 262; Gaps 41;

QY 21 IDGKAQTKTENITPD--EYKRE---GNAEQIVIKITDQGVTS-----HGDIHYH 68
Db 193 LEKQSLIKDFDENDEYELSEKNSDQSSPSIMI-LSDEYAEAGALQDVSNDYAE 251
QY 69 YNGKVPYDAIISBELMKD-----PNY-----KLKDE-DIVNEVK----- 102
Db 252 EEOVERKNIGQOANVENATQISSDSSSGYQSEGVENELEDDIDVESDAKDSOGA 311
QY 103 GGVIVKDGKYVYLKDAADHNVTRKEINRQKQHSQHREGGTPRNDGAVALARSOGR 162

```

```

Db 312 EGTEHSDVFSKYM---QPRDTNTKIPVTEKYESDEHKVHQ-----R 349
QY 163 YTTDDGYIFNASDII-----EDTDGAYIVPHGDHYHVIIPKNELASALAAAFLSG 214
Db 350 YSEDGAFDFGSVNI SVDDDESEDEESQAESYANAEVYHH-NEHLEDDKEL--IEDIES- 405
QY 215 RGNLSNSTYRONSNDTSRTNWPSVSNPGTTNTNTNNSNTNSQA-----S 262
Db 406 ---SDSESQAESQESQEDDFEYKMKNEKSTSETENTSESRDQGFADKAYTKNKVEQ 461
QY 263 QSNIDID---SLLKQLY---KLPLSRHVESDGLVFDPA-----QITSRT 300
Db 462 QENDEEPEKDDLIIRSLDKNPHGNKSEYSENVLENE---TDPALVRENOINDEGYD 518
QY 301 ARGVAVPHGDHYH-----FIPYSQ-----MGELEERIIARIIPLY--RSNHW 341
Db 519 VTGKSVESDLHEHSPDNLVLAARAMLQFQOSRNSNCPCQKEEQVSE---SVLHGNSGN 574
QY 342 PDSRPEQSPQPTPEPSPQAPNPKIDSN--SSLVSQLVKVGEGVFEKGIISRYVF 399
Db 575 LSGRSLDESEEQIPLKDFGTENNNNLKTDRGLSSSVIEVEKVSE----- 620
QY 400 AKDLPSQTVKML-----ESKLSKQESVSHTL-----TAKKE 430
Db 621 -KKLDGSTKELVPLSTDTTINNSLGNEDSIYSLDDADALISEMLTDPVPLMEIKTTPKY 679
QY 431 NVAPRDOEFYDKAY--NLLTEAHKALFXNKGNSDFQALDKLLRLNDESTNKEKLVDDL 488
Db 680 EVWISSEVSYSTSYEDNTVAMPQVEVTSPPMNDPFNS-----LNDDYEKKHLLAKST 732
QY 489 LAFLAP-ITHEP-----RLGKPNESQIETDEVR-----AQLADKYTSDGYIFDHD 536
Db 733 LAALAPFTKKDAEFVAGVTKSLTSTSGHTNIFHTSKETKQVSDLDSTENVTFF--ENE 791
QY 537 IISDE--GDAYVTPHMGHSHWIGKDSLSDKKAQAAQYTKKGIILPPSPDADVKANPTGD 594
Db 792 NTGDNKNQSKNFPFVANSSTDKSTEDNTDEKYFSAINVT-----NVTGD 835
QY 595 SAAIYNRVKGEKRIPLVRLPYMVVHTV-----EVKNGNLIIPHKD--HYHNKIF 642
Db 836 SSCS-----DIETASNVENLRYCEKDMNEAMSSGDECVKQNDGSKTQISF 884
QY 643 AWFDDHTYKAPNGVTLDELATIKYVVEHPDERPHSDGNGWGNASEHVLGKKD--HSDPDN 700
Db 885 STGSPDNFQSNNDT---EFSSTKYKVRNSDLEDESLKKELTKARVVDVKLDEESEDYS 941
QY 701 KNFKADEEPEVEETPAEPE-----VPQVETEKVAQKAEAEVLLAKVTD 744
Db 942 EQDYADPEPGNDEGSNENIVKGTGKOTLGIVEPENKVNKVEETLFEANVSS 996

RESULT 6
SC16_YEAST STANDARD; PRT; 2195 AA.
AC P48415; Q02822;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Multidomain vesicle coat protein.
GN SEC16 OR YPL085W OR LPF1W
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017704; PubMed=7593161;
RA Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that
RT interacts with Sec23p."
RL J. Cell Biol. 131:311-324 (1995).
RN [2]
RP SEQUENCE FROM N.A.

```


RC STRAIN=S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansonge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Borstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friese J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hallier L.W.,
RA Hunkle-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Pohl F.X., Pohl T.M., Purnelle D., Schafer M., Schaefer M.,
RA Scherens B., Schramm S., Schroeder M., Sidic A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.M., Zollner A., Vo D.H., Hani J.,
RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI";
RL Nature 387:103-105 (1997).
CC -!- FUNCTION: Involved in the budding of transport vesicle from the
CC endoplasmic reticulum. The C-terminal interacts with SEC23 and
CC with the cytosolic domain of SED4. Could therefore be a
CC constituent of copII vesicle coat. N-terminal overexpression
CC causes a lethal secretion defect.
CC -!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
CC WHICH BUD FROM IT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U23819; AAC49088.1; -;
DR EMBL; U41849; AAB68254.1; -;
DR PIR; S61103; S61103;
DR Gene; 144067; -;
DR SGD; S0006006; SEC16.
DR GO; GO:0030127; C:copII vesicle coat; IDA.
DR GO; GO:0005789; C:endoplasmic reticulum membrane; IDA.
DR GO; GO:0005198; F:structural molecule activity; IPI.
DR GO; GO:0005914; P:autophagy; IMP.
DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094
FT MUTAGEN 1059 1059 LYS-RICH.
FT L-S: IN SEC16-4; TS ACCUMULATION OF ER
FT MEMBRANES
FT L-P: IN SEC16-3; TS ACCUMULATION OF ER
FT MEMBRANES.
FT L-P: IN SEC16-2; TS ACCUMULATION OF ER
FT MEMBRANES.
FT W-R: IN SEC16-1; TS ACCUMULATION OF ER
FT MEMBRANES.
FT MISSING (IN REF. 1).
FT I -> F (IN REF. 1).
FT CONFLICT 522 522
FT CONFLICT 560 560
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;
Query Match 3.8%; Score 157; DB 1; Length 2195;
Best Local Similarity 19.7%; Pred. No. 1.8;
Matches 124; Conservative 109; Mismatches 241; Indels 156; Gaps 31;
QY 188 HGDHYHYPKNEIASAEALAEAFISGRGNLSNRTYRQNSDNTSYNWPVSNGT 247
Db 29 HSEPLELPSTINSS-----FNDDSVNRTESDIASKD-VPPVS-----S 68
QY 248 NTNYS-----NNSNTNSQASQNDI-----DSLKLKLYKLPL 279
Db 69 STNISSPANETOLEPTDQELHKLNDQDIDITADGNDPNSIVHDSVITQT-KPAM 127
QY 280 SQHVSDGLVDFDAQITSR-----TARGVAVPHGDHYHFFIPYSQMSE----- 322

Db 128 SQYEET-----AAHLSRRNPSLDVWAGELHNNNEHTQKIAVSAAVEEDSFNEEGENHD 181
QY 323 --LEERIRIPLRYSRHHWPDSPRQPSQPTPEPSPQOPAPNLKIDNSLSVSLV 380
Db 182 SIIISLNDATPSQY--NHLPSPD-----GNLSPELSSGDTPTTHNPLGTVDN----- 228
QY 381 KRVGGYVFEEKGISRYVFAKD-LPSETVKNLESKLSQESVSHLTITAKENVAPRDQEF 439
Db 229 -EINDEVCNDKEIS--LNANNVLPDELSEKDEERLKLETHVS--TEEKQDIA--DQET 281
QY 440 YDKAYNLLTEAHKALFXNKNRSD--FOALDKLLERLNDESTNKELVDLLAFIPIH 497
Db 282 AENLFTSSTPESENKIRNSGDDTSMFQ-----DDESQKVPWEEDVKDF---H 328
QY 498 PRLKQPNISQIYTTEDVR-----IAQLADKYTTSDGYIPDE---HDIISDEGDAYVTP 548
Db 329 NENTNNTQESAPNTDORDKGYEGNALKKSECTAADERSYSEETSEDIFGHDKQVVE- 387
QY 549 HMGSHWIGKDSLDEKVAQAQYTKKGILPPSPDADV--KANPTGSAAIYNRVKG 606
Db 388 --GQNDFTGKNTEENSQKLMGSGNHK---LPLSAEADIIERPKDIDQDAEDLFTQSSGD 441
QY 607 KRIPVRLPYMVEHTVVEVNGMLIIPHKDHYNI-----KFAW----FDDHTYKAPN 654
Db 442 -----LGEVLPWESTDKNADVTSQEKHEDLFAASGNDEKLPMEVSDGEVSSGKTEN 494
QY 655 G-YTLEDLPATIKY-YVEHPDERPHSNDGWNASHVLGKDKHSDSEPNKFNKADEPVEE 712
Db 495 SWQSTTEKIAEQKFLENDDLLDDDDSLASSE-----EEDTVPTNTDNTLTSKVEE 550
QY 713 TPAEPPEVPOVETEKVEAQLKEAEVLLAKVT 742
Db 551 KKAARYKPIEB---EAGMRQEQVHFTNTT 577
RESULT 7
MAPX DROME STANDARD; PRT; 1185 AA.
AC P23226; Q9V9S1;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 205 kDa microtubule-associated protein.
GN MAP205 OR CG1483.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP MEDLINE=91115949; PubMed=1703540;
RX Irminger-Finger I., Laymon R.A., Goldstein L.S.B.;
RA "Analysis of the primary sequence and microtubule-binding region of
RT the Drosophila 205K MAP";
RL J. Cell Biol. 111:2563-2572 (1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM B3).
RX STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.N., Cawley S., Dahlke C., Davenport L.B., Davies P.,

Qy	122	HADNVRTKEEI-NRQKQ-----BHSQREGGTPRNDGAVALARSGRYTTDDGYIF	171
Db	4	HEDNALDNYLQNRLAESLQICGAGEHNPHLADTGGNCAPGAPSK-----	52
Qy	172	NASDIIEDTGDAYIVPHGDHYHYIPKNLSASELAAAE-----AFLSGRGNLSNS	221
Db	53	---SDEVGDEED-----EEMKYYI---HEVRQSEKIQEKLPLTKETGNGFGPGR-----	95
Qy	222	RTYRRONSDDTNRTNW-----VPSVSNPCTTNTNTNSNNTNS-----QASQSD	266
Db	96	-----DSDNQVHGNGAAAVPNLVEEDVEVTKNDGDFSTNTTTTDEVVARQEPNQ	149
Qy	267	IDSLLKQLYKLPLSORHVESDGLVDFDPAQITSTRARGVAVPHG-----DHYHF	314
Db	150	LPEQLQ-----QQQIESQGVHEDPRQEDEHSSVATTGYTSSLENNSSPLDQEV	202
Qy	315	IPYSQ-----MSELEERTARIIPLYRNNHWVDSRPRQPSQOPTPE	356
Db	203	VMVAQTVGQQLVDFDNKENSIFYKNLEENHSQJNP---NAVAFVPGVGSOSSSPLPAE	259
Qy	357	PSFGPQAPANLKIIDSNSLSVLQVRKVGEGYVFEKGISRYFAKOLPSETVKNLESKL	415
Db	260	DELPGVQPRFPLPGTGLDDLVAESPCK---EFARINMDGIA-----VPDREDFEADM	310
Qy	416	SKQESVSHTLTAKENVAPRQDEYDKAYNLLTEAHKALFNKNGNSDFQALDKLLERLN	475

Db	311	R----	PHELEQSDTFGAGHLEMQ-----	LLNGTGTADQAALRDVLDP	351
Qy	476	DESTNKEKLVDDLLAFIAPITTHPERLCKPNSQIF-----	YTEDEVRIQLADKY	524	

Db 352 ETSVDME-LPLQOVNDADIMKQSIYAEHNSIEDILNSVQLPIQTCDKELHVEEKE 410

411	HVSKSPSTBELQPSD-----FPNQESHTLFNNTQDPMQASFYLEHTSQKAQEGCQE	464
Db		
525	TTSDGFIYFEBHDIIISDEGDAYVTPHMGHSHWIKDLSLSDKEKA-----AAAYTKE	575
QY		
576	KGTLPPSPADVKNKNT--GDSAAAYNRYKGEKRIPLVRL-----	614
QY		
465	QMQL-PAECSDIPADQSLDLDTSAP---QLSSRADSPVAKLELESQAQAGVITDTPSPSS	520
Db		
615	---PYWVEHTVVKNGNLIIPHKDHYHNIKFAWFDHDTYKAP---NGYTLIEDLFATIKY	667
QY		

[illegible]

628	DB	MDVVGMP:IPASEDF	FAICDKVASKSNEVEDHRS	EQQAFVKEELLHPG	VDWAQENLGT	687
768	QY	NNSIMAAEKKLLALIKG	SNP-SVSKEKI	795		
688	DB	EKNFVVEERLPISV	DEIPLSSASKEKL	716		

RESULT 8

ANK2_HUMAN	
IID	ANK2_HUMAN
AC	Q01484; Q01485;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN	ANK2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Mammalia; Primates; Hominidae; Homo

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N A (ISOFORMS 1, 2 AND 3)
RP

RC
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
TISSUE=Brain stem;

RX	MEDLINE-91302466; PubMed-1830053;
RA	Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT	"Isolation and characterization of cDNAs encoding human brain
RT	ankyrins reveal a family of alternatively spliced genes.";
J.	J. Cell Biol. 114:241-253(1991).
[2]	
RN	REVISIONS.
RP	Carpenter S.;
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[3]	
RN	SEQUENCE FROM N.A. (ISOFORM 1).
RC	TISSUE=Brain stem;
TX	MEDLINE-94075409; PubMed-8253844;
RT	Chan W., Kordeli E., Bennett V.;
RA	"440-kD ankyrinB: structure of the major developmentally regulated
RT	domain and selective localization in unmyelinated axons.";
J.	J. Cell Biol. 123:1463-1473(1993).
[4]	
RN	SEQUENCE OF 463-495 FROM N.A.
TX	MEDLINE-92009921; PubMed-1833308;
RA	Tse W.T., Menninger J.C., Yang-Peng T.L., Francke U., Sahr K.E.,
RA	Lux S.E., Ward D.C., Forget B.G.;
RT	"Isolation and chromosomal localization of a novel nonerythroid
RT	ankyrin gene.";
RL	Genomics 10:858-866(1991).
CC	-I- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC	elements. Also bind to cytoskeletal proteins.
CC	-I- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=3;
CC	Name=1;
CC	IsoId-Q01484-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId-Q01484-2; Sequence=VSP_000267, VSP_000268;
CC	Name=3;
CC	IsoId-Q01484-3; Sequence=VSP_000268;
CC	-I- TISSUE SPECIFICITY: Plasma membrane of neurons as well as glial
CC	cells throughout the brain.
CC	-I- PTM: phosphorylated at multiple sites by different protein kinases
CC	and each phosphorylation event regulates the protein's structure
CC	and function (Potential)
CC	-I- SIMILARITY: Contains 23 ANK repeats.
CC	-I- SIMILARITY: Contains 1 death domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL	X56957; CAA40278.1; --
DR	EMBL; X56958; CAA40279.2; --
DR	EMBL; Z26634; CAB42644.1; --
DR	EMBL; M37123; AAA62828.1; --
DR	PIR; S37431; S37431.
DR	HSSP; P42771; 1DC21.
DR	Genew; HGNC:493; ANK2.
DR	MIM; 106410; --
DR	InterPro; IPR002110; ANK.
DR	InterPro; IPR000488; Death.
DR	InterPro; IPR000906; ZU5.
DR	Pfam; PF00023; ank; 24.
DR	Pfam; PF00531; death; 1.
DR	Pfam; PF00791; ZU5; 1.
DR	PRINTS; PR01415; ANKYRIN.
DR	SMART; SM00248; ANK_22.
DR	SMART; SM00005; DEATH; 1.
DR	SMART; SM00218; ZU5; 1.
DR	PROSITE; PS50088; ANK REPEAT; 20.
DR	PROSITE; PS50297; ANK REP REGION; 1.
DR	PROSITE; PS50017; DEATH DOMAIN; 1.
RW	Croskeleton: Alternative splicing. Repeat: ANK repeat;

Db 3194 TVQTGDIPPLS--GVKQISCPDSSSEPAVQVQL--DFSTLRSVY-----SD----- 3235
 QY 292 DPAQITSTARGVAVPHGDHVFHIFPYSQMSLEERARIILPIRYSNHWVDSRPEQP-S 350
 Db 3236 -----RQDDSPDSS-----PEBQSVIE-----IPTAPMENVPFTESKSIPIVR 3274
 QY 351 POPTPEPSPGPOPAPNLKIDNSSLSVLSQVLRVKGEGY---VFEEKGISRYVFAKDLPSRT 407
 Db 3275 TMTPTSTPAP-----PSAEVSS-----VSEDFLSSVDDEENKADAPKSKLPVKV 3319
 QY 408 -VKNLESLSKQESVSHTLTAKKNVAPRDQEFYDKAYNLLTEAHKALFLXNKNRGN-SDFQ 465
 Db 3320 PLQVREQQLSDLD-----TSVQKTVAPQGDMAIAPD-----NRKSSESAS 3362
 QY 466 ALD-----KLLERLNDESTNKEKLVDDLLAFIAP-----PVKSRSTTS 3418
 Db 3363 SLDSKTKCPVTRSTETETESRERAELESEEGATRPKILTSRL-----PVKSRSTTS 3418
 QY 496 -----THPERLGKN-----SQIBYTE-----DEVRIAOLADKYTSDGVIPEHDDIISDE 541
 Db 3419 SCRGTSPTKESKEHFFDLYNSIEFFFEISDEA--SKLVRLTQSE-----REQEIVSDD 3472
 QY 542 GDVYTPHMGSHWIGKSLDSKEXKAAQAQYKTEKGILPPSPDADVKANPTGDSAAAIYN 601
 Db 3473 -----EPEVQVETEKVQAQKAEVLL-----AKVTDSS 745
 QY 602 RVKGEKRIPLVRLPMVHEVTVKXNGNLIIPHKHQHYHNKFAWDDHTYKAPNGYTLDEL 661
 Db 3479 EVSVIENLP-----PVEHSHVP-----ED1 3499
 QY 662 FAT-----IKYVVEH--PDRPHSNDGWNASEHVLGKDKHSEDPNKNFKADEPVEET 713
 Db 3500 FDRPIWDESJETLIERIPDENGH-----DHAEDP-----ODQERIEER 3539
 QY 714 PA-----EPEVQVETEKVQAQKAEVLL-----AKVTDSS 745
 Db 3540 LAYIADHLGFSWTELARELDPTEQIHQIRIENPNSLDQSQYLLKTVLWRDQGHATDTN 3599
 QY 746 L-----KANA-----TETL-----AGLRNLTLOIMDNNISMAEKLALL 782
 Db 3600 LVECLTKINRMDIVHLMETNTEPLQERISHSYAEIEQTITLDHSEGFSLQOE-BLCTAQH 3658
 QY 783 KGSNPSSVSKE 793
 Db 3659 KQKEEQAVSKE 3669

RESULT 9

Y373 BOVIN
 ID Y373 BOVIN STANDARD; PRT; 1453 AA.
 AC Q9TU23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0373.
 GN KIAA0373.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Trachea;
 RA Jovov B., Ripoll P.J., Benos D.J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF176816; AAF00990.1; -
 KW Hypothetical protein; Coiled Coil.
 FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF4139 CRC64;

Query Match 3.7%; Score 154; DB 1; Length 1453;
 Best Local Similarity 18.5%; Pred. No. 1.5;
 Matches 171; Conservative 124; Mismatches 329; Indels 298; Gaps 40;

QY 7 YQARTVKNNRVSYIDGKQA-----TQKENTL---PD-----EVSKEGGINAQSI 49
 Db 486 WDQREVELEROLDVDFDQOQSEILREAOKEFEATGSMPPDPSLPINQLEIALRK--IKENI 543
 QY 50 VIKITDQGVYVSHGDHVVYNGKVPYDAIIIEELMKDPNPKDKEDI-----VNEVKG 103
 Db 544 RILLETQATCS-----LEEKREKESALRLAEENILSRDKVINEUR- 585
 QY 104 GYIVKVDGKYVYLKDAHADNVRTKEINRQKQHSQREGGTPRNDGAVALARSGRY 163
 Db 586 -----LRLPATAEQEKILAEFSRKEVEPKSHHTLKLAHQ-----TIANQOARL 628
 QY 164 TTDGDIYNASDIIEDTGD---AYIVPHGDHYHYI-PKNELSASELAABAFSLGRNLS 219
 Db 629 NQKEEVLLKYOHLEKAREEQREIVKKBHEELHTLHRKLELQA----- 671
 QY 220 NSRTYRRONSNT-----SRTNW---VPSVSNPTGTTNTNTSNNN-TNSQASNDIDSL-- 270
 Db 672 -----DNSLSKFKETAWDLIKQSPVTPVFNKHFIRLAEMEQTVAEQDSSLSLVI 721
 QY 271 LKQI-----YKLPQSQRH-----VESDGLVDFPAQITSR 299
 Db 722 KLVQVQDLERQKEITELKKEFENMKLRLQENHAEVKKIKAEVEDLRCLLVQSQESQ 781
 QY 300 TARGVAVPHGDHYHIFPYSQMSLEERARIILPIRYRS-----NHWVPDSRPEQSPQTP 355
 Db 782 SLKSELOTQKANSRAPTTTMRNLVERLKSQALKEKQKALSALLLEAEAMTAABER 841
 QY 356 EPSGPGOPAPNLKIDS-----NSSLSVQLVRKVGEGYVFEKGISRYVFAKDLSETVKN 410
 Db 842 IISWTSQKEANLVQOIVDRHTKELSQI-----EDL-NENILK 879
 QY 411 LESKLSKQESVSHTLTAKKNVAPRDQEFYD--KAYNLLTEAHKALFXNKNRGNDSDFQALD 468
 Db 880 LKEALTKSKNRENTLT---DNLDLITNELQNKQKAYG-----KVLREKDAVDQENNELK 930
 QY 469 KLLERLNDE-----STNKEKLVDDLLAFIAPITHPERLGKPNQSI EYTEDEVRIQALAD 522
 Db 931 RQIKRLTSGLGKPLIDNKQSLIEEL-----QKKIKLESQLEKRVKDEAEKMPKE 981
 QY 523 KYTTSDGYIFDE-----HOIISDEGDAYVTPHMGSHWIGK--SLSDKQKVA 568
 Db 982 KSAREEVIRWEEGKKMQTKIEGRNKLKEGEVYI---LTQLTTLTKDLIPAKADKELT 1038
 QY 569 AQAYTEKGLPPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPMVHEVTVKXNGN 628
 Db 1039 LQRLKLTGTL-----TVQVMAA-RVLESEKSELE-----ELKRN 1072
 QY 629 LIIPKDHVHNKFAWDDHTYKAPNGYTLDEL FATIKYVVE--HPPERPHSNDGWNAS 686
 Db 1073 LDLENDISYMRSHQA-----LPRDSVIEDLHQLNKYQEKLEHALEKQLSKDAYSRPS 1124
 QY 687 EHVLGKDKH-----SEDPNKNFKADEPVE-----ETPAEPEVQVET----- 724
 Db 1125 TSGIDSDHVVQREQLQRENKLSSNELKFOLEQAKNDLPRLKNQVRLKEMCEFLKK 1184
 QY 725 EKVEAQKAEV-----LAKVTDSSLIKAN-----ATETLAG 756
 Db 1185 EKAEVERKLGRVGRSGRSGKTIPELEKTIQMKVVEKQRENEQLKASGILTSEKMAN 1244
 QY 757 LRNNLTLOIMDNNISMAEKL 778

```

1245 IE-----MENEKLAELK 1259
DB

RESULT 10
Y373 HUMAN
ID Y373 HUMAN STANDARD; PRT; 1539 AA.
AC O15078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB02371; BAA20828.1; -
KW Hypothetical protein; Coiled coil.
FT DOMAIN 18 1514
FT SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;
SQ
Query Match 3.7%; Score 153.5; DB 1; Length 1539;
Best Local Similarity 19.2%; Pred. No. 1.7;
Matches 200; Conservative 145; Mismatches 334; Indels 361; Gaps 53;
QY 3 ELGLYQARTVKNRVSYIDGKATQKTEN-----LTPDEVSKREGINAEQIVIKITD 55
DB 306 EQALYYAR-LEGRNRAHL--RQTQSLRRQFSGALPLAQQKFKSKTMIQLQNDKLIKIM- 361
QY 56 QGVVTSHGHHYNGKVPYDAIIS--EELL--MKDP-----NYKLKDEDI-VNEVKG 103
DB 362 QEMKNSQOHRNMENKTLEWELKGLGELISTLKDTKGAQKVINHWMTIELRLQELKL 421
QY 104 GYVIVKDGKYYVYLKDAHADNVRT---KEENRQKQHSQHREGTTPRNDGAVALAR 159
DB 422 NRELVDKKEILNNII-SEVERTTSSLEEIVQONKPFEEERQMAWDRE---VDLERQ 477
QY 160 QGRYTTDDGYFNASDIED-TGDATVPHGDHVVHPKNEI----- 200
DB 478 LDIFDRQNEILNAAQKFEATGS---IP--DPSLPLP-NOLIEALRKIKENIRILETR 531
QY 201 -----SASELAFAAEFLSGRGLNSRTYRRQNSDNTSRTNWPVSVS-----NP 244
DB 532 ATCKSLEELKKEKESALRLAEQNLSDKVINB--LRLRLPATAEREKILAEIAGKXEMEP 589
QY 245 GTTNT-NTSNNTNGQASQNSDIDSLKLYKPLPSORHVESDGLVDFDPAQITSTRARG 303
DB 590 KSHHTLKIAHQTIANMQA-RLNQKEEVLYKKYQRLLEKAREEQ-----RE 632
QY 304 VAVPHGDHGHYFIPYSQMSLEERARIIPURYSRHNWVPDSRPBPQSPQPT----- 354
DB 633 IVKHEEDLHLHLRLQADSSLN-----KFKQTAW--DLMQSQFPFVTKNHFPLAE 685

```

```

355 PEPSPGPAPNLKIDSNLSVLSQVR-----KVGEGYVFE 390
DB MEQTVAEQD-----DSLSSLLVLKLVKSQDLEROREITELVKKEFENIKLQIQENHDE 739
QY 391 EKGISRYVFAKDL-----PSETVKNLESKL----- 415
DB 740 VKKVAEV--EDLKYLLDQSKESQCLKSEIQAKSEANSRAPTTTMRNLVERKLSQALAK 797
QY 416 -SKQESVSHIL-----TAKKE--NVAPRDQBFYDK----- 442
DB 798 EKQKALSRALLELEPAEMTAAAEERIIISATSKAEHLNV-----QOIVDRHTRELKTQVED 853
QY 443 -AYNLLTEAHKALFXNKR-----NSDFQ--AUDKLL-----ERLND----- 477
DB 854 LNNELL-KLKEALKTSKNRENSLTNDLNDLNNELQKQKAYNKILREKEEIDQENDELKR 912
QY 478 -----STNKEKLVDDLLAFAPITPHERLGKPNQSOIYETDEVRILQADK 523
DB 913 QIKRLTSGLGKPLTDNKKSLIEEL-----QRKVKLENQLEGKVEEVDLKPMEK 963
QY 524 YTTSDGYIFDE-----HDIISDEGDAYVTPHMGHSHWIGK--SLSDKEKVA 569
DB 964 NAKELIRWEEGKKWQAKIEGIRNKLKKEGEVFTLTQKQNTL--KDLFAKADKEKLT 1020
QY 570 QAYTKEKGLPPSPDADVKANPTGSAAIYNRVKGKRIPLVRLPYMVEHTVEVKNGL 529
DB 1021 QRKLKTTGM-----TVQVLGI-RALESEKELE-----ELKKNL 1054
QY 630 IIPKDHVHNIFAFDDHTYKAPNGYTLDELFIATIKYVVE--HPDERPHSDMGWGNASE 687
DB 1055 DLNDILYMRHQA-----LPDSDVVEDLHLQNRVLOEKHLALEKQFSKDTYSKPSI 1106
QY 688 HVLGKKKH-----SEDPNKNFKADEPVE-----ETPAEPVEPVQVET-----E 725
DB 1107 SGIESDDHCHQREQLQKLENKLSENIELKFQLEQANKDLPRLNQVDRDKEMCEFLKKE 1166
QY 726 KVAQAKAEV-----LLAKVDTSSLKANATETLAGRNNTLQIM- 766
DB 1167 KAEVQRKLGHRGSGSGSKTIPELEKTIGLMKKWEKVQREN--EOLKKASGILTSKWA 1224
QY 767 ----DNNSIMAEAEKLLALL 782
DB 1225 NIEQENEKUKAELEKLAHL 1244
QY
DB
RESULT 11
ID RBP2_PLAVB STANDARD; PRT; 2867 AA.
AC Q00799; Q9N2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10818229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares
RT structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa
RT rhoptry protein family.";
RL Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -|- FUNCTION: involved in reticulocyte adhesion. Specifically binds to

```

human reticulocyte cells.
 --!- SUBCELLULAR LOCATION: Membrane-bound (Probable).

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF184623; AAF76525.1; --
 DR HSSP; P03069; 1GCM.
 KW Malaria; Receptor; Signal; Transmembrane; Repeat.
 FT SIGNAL 1 21
 FT CHAIN 22 2867
 FT DOMAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.
 FT TRANSMEM 2806 2826 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 2827 2867 POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 44 133 ASN-RICH.
 FT DOMAIN 560 758 LYS-RICH.
 FT DOMAIN 1112 1285 LYS-RICH.
 FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.
 FT REPEAT 2758 2761 1.
 FT REPEAT 2762 2765 2.
 FT REPEAT 2766 2769 3.
 FT REPEAT 2770 2773 4.
 FT REPEAT 2774 2777 5.
 FT REPEAT 2778 2781 6.
 FT REPEAT 2782 2785 7.
 SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

 Query Match 3.7%; Score 153.5; DB 1; Length 2867;
 Best Local Similarity 18.5%; Pred No. 4.1;
 Matches 167; Conservative 143; Mismatches 348; Indels 247; Gaps 39;

 QY 12 VKENRV-SYIDGKQATQKTENLTPDEVSKR-----EGINAEQIVIKITDQGYVTS--- 61
 DB 663 ISKNDIEQKYI-----EKIEKLINDAPSGDKFTTEKTNLQNKVKIIDFHKEDLQLL 716
 QY 62 -----HGDHYHYN-----GKVPYDAI-----ISELLMKDPNY 90
 DB 717 LNSLSKPYEEHQKLYNEASTIEKIKDLHQKTKBEYEKLEKMFNSFGQILDKLNTLNDL 776
 QY 91 KLKDEDIVNEKGGYVIVKVDGKYVYVYLKDAADANVRT-----KEENRQ 135
 DB 777 KTLKKNIVEB-QTYNINKWSDSLTNL--TAEVDNLSALDGYADELTQYKNRNER 833
 QY 136 KQEH---SQHREGTTP-----RNDGVALARSQGRYTTD---DGYIFNASDIIE-- 178
 DB 834 KEKFLSTLKEQEDDIPDGKNIYEEYNNHKNVNMVKNKEHKISSDINQCNENIIEKKNLETF 893
 QY 179 ----DTGDAYIVPGDHYHYPK-----NELSASELAABAFUSGRGNLSNRYR-RQ 227
 DB 894 NTLVQTLDAHTGKKQKVDHLLQKFKTNLEKLNLELESFGKSLNGSASTTNKQIENIRK 953
 QY 228 NSDNTSRTNWVPSVSNPQTNTNTNNSNTSQAQSDNDISLLKQYKLPQSQRHVESD 287
 DB 954 NIDTIKSLNF-----AKNSSESKLSLENIKNKADLIKLDQHTQOEIEK- 998
 QY 288 GLVPDPAQITSRTARGVAVPHGDHYHFPYSQMSSELEARIIPLRYSNHWHPDSRPE 347
 DB 999 -----HTFIENEEMSP-----LSVIKKEKN----- 1019
 QY 348 QPSQPTPEPSGPQAPNPKIDNSLSVQLVKVGGYVFEKGSRYVFAKDLPSSET 407
 DB 1020 -----RVESDMSSEELIKQLNTKINALEIYNNKSKDRFGDDETNLEE 1061
 QY 408 VKNLESKLSKQESVSHLTAKENAVPRDQEFYDKAYN-LLTEAHK---ALFYXNKGNSD 463
 DB 1062 LDFPKQCCQAQBIKKLTT-----NPNVLDNGINVIIEQHEKVIILSENHITKED 1113
 QY 464 FOALDKLLERLNDESTNKEKLVDDLLAFAPITHPERLKGPNQIEYTED---EVRIAQL 520

DB 1114 KKINEKIQNVNSLNMETKL-----GLLKINEDIKNSRDTTICKSKIOEF 1158
 QY 521 ADKTYTSDGYIFDEHDIISDEGDAYVTVMGHSHWIGKDSLSDEKVAQAAYTEKKGILP 580
 DB 1159 EKKVQT-----IFGSDIVANKKIDA-----IKKEHDVNKDEF-DKEKVDTSFDEKKKSIE 1208
 QY 581 PSPDADVAVKANTGUSAAAIYNRVKGKRIPLVRLPY---MVEHTVEVKNGL-----IIPH 633
 DB 1209 KAYE---KXGNTLKELEKMDDEKNEKEVEEAQIQYKRIFFIDHDVNLNMDNEVEKSKIYME 1265
 QY 634 KDHYHNKIFAWFDHHTYKAPNGYTYLEDLFIATIKYVVEHPDERPHSN---DGMGNASEHVL 690
 DB 1266 KIELYKKEI-----DEIKQKTYEYKQGD---TSNFYITEQYNSATQSKAKIEQFINIATTKK 1319
 QY 691 GKQKHSDEPNKNFKADEPVEETPAEPVQVQVETEKVEAQLKE-----AEVLLAKVTDSLS 746
 DB 1320 GTSDTSQDINEL-----ESIKEEVHKNLQLVQESNMSMEERKQILSMKDLLILNSETIA 1375
 QY 747 K--ANATETLAGLRNNL-----TLQIMDNNSIMAEAKL-----LALLKGSNPFSSVSK 792
 DB 1376 KEISNNTQNALGFRENAKTKLNTDELLQORVAAMIEEAKAHKNNDIDIALEDAQIDTEVSK 1435
 QY 793 -EKIN 796
 DB 1436 IEQIN 1440

 RESULT 12
 YOH8 YEAST
 ID YOH8 YEAST STANDARD; PRT; 1176 AA.
 AC Q08236;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 131.4 kDa protein in REX4-ATP19 intergenic region.
 GN YOL078W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97321807; PubMed=9178509;
 RA Tzermita M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast
 chromosome XV reveals eight known genes and ten new open reading
 frames including homologues of ABC transporters, inositol
 phosphatases and human expressed sequence tags";
 RL Yeast 13:583-589(1997).
 CC -1- SIMILARITY: BELONGS TO THE SIN1 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; Z74820; CAA99089.1; --
 DR EMBL; Z74819; CAA99088.1; --
 DR PIR; S66771; S66771.
 DR Germonline; 143500; --
 DR SGD; S0005438; YOL078W.
 DR GO; GO:0005737; C:cytoplasm; IPI.
 DR GO; GO:0001558; P:regulation of cell growth; IPI.
 DR InterPro; IPR008828; SIN1.
 DR Pfam; PF05422; SIN1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1176 AA; 131378 MW; ED2B30D014EB5652 CRC64;

 Query Match 3.6%; Score 150.5; DB 1; Length 1176;

QY 76 -----DAITSEELMKDPNPKLDEDIVNEVKGGYVIKVDGKYYVYKLDAAAHADNVRTXE 130
: : || | : : || : || | : || :
: : || | : : || : || | : || :

```

Best Local Similarity 19.4%; Pred. No. 1.9;
Matches 138; Conservative 87; Mismatches 273; Indels 213; Gaps 30;

QY      29 KTENLTPDEVSKRGINAEQIVIKITDGGVTS-----HGDHVHYNGKVY----- 75
          |||         |||
Db       202 EAEMLYMGEVERMDGYESPAK-DSQSDISIGACLDGIFVKXGNRPVPVFRHWDIA 260
          |||||        |||
QY      76 -----DAIIEELLMKDPNYKLDEDIVNEVGKGYVIKVDGKYVYLKDAHAONVRTKE 130
          :   :   |   :   :   :   :   :   :   :   :   :   :   :   :

```

Db 261 NMHNKSFPALELANKEETIQOTEDMET-----AKYVWRLCVARHKFYRLNQC 309

QY 131 EINROKQHSQREGTSPRNDGAVALSQ-----GRYTTDDGYIFNASDII 177

Db 310 NLOQAAALNSVRRGSSR---MSLPKPQPYAMPFPQALHYNGHYT-----BPF 355

QY 178 EDTGDAYIVPHGDHYHYTPKNELSASAAAAEFLSGRGNLSNRTY--RRONSNTSRT 235

Db 356 ASSQDNVFPVKNKGYCHQSLSLDTQID-----LSGR--IRNGSVYSAHSTNSLNTQ 408

QY 236 NWVPS--VSNPQTTWNTNSNTNSQASQ-----SNDIDSLKQLYK-LPLSORH 283

Db 409 YLOPSPMSNPSIPGSDVMRPDIYPSHRHSALIPPSYRPTDYESVMKRLNRMVHADRH 468

QY 284 VES-----DGLVDFDAQITSTRARGVAVPHGDHYF-----IPYSQ 319

Db 469 SHSLNLSIGSYAYSRPDALVYSOPEI--REHPLHASPQSAHYFNLNYSFHSQAPYPY 526

QY 320 MSELERARIIPLYRSNHWVPDSRPE-----QPSQPT-----PEPSGPQ 362

Db 527 PVERRPVVGAV-----SVPELTNVQLQAQDYAPAPNIMRTQVYRPPPPYPR 573

QY 363 PA---PNLK-----IDNSNSIVSQLVRKGVGYVEEKGISRYVFAKDLPSSETVKNLESK 414

Db 574 PANSTPDLRHLIYSSSNPDLITRRVHSVQ--TFQEDSLPVAHSLQVSEBPLTAARHAH 631

QY 415 LSKQESV-----SHTLAKKENVAPR-----DOEFYDKAYNLLTEAH- 451

Db 632 LQKNSNLSIAGLTGFEGLRKEETMSASAADVAPRTFSAGSQSVFSDKVQBGTEBQ 691

QY 452 KALFANXKGRNSDFQAL---DKLLRLNDESTNKEKLVDDLE-----AFLAP 494

Db 692 SGGYSHKSLSDATMLIHSSEDELEDDSSREHAVSEPLTAAPFSQEQQLNYPCASVTP 751

QY 495 ITHPERLGKNSQIETVEDEVRIALQADKYTTSQYIPDEHDIISDEGDAYVTPHMGSH 554

Db 752 VTGFLHIFEPKSHV--TEPEKRAKDIPVHLVME-----THQPRRH 790

QY 555 WIGKDSLSDEKVAQAAYTKEKGILPPSPDADVANKPTGDSAAALYNEVK 605

Db 791 GLLTFSMESDLTTSGRYARRDSL-----KKRPVSDLLSGKNTVEG 833

RESULT 14

BD1 PLAVB
ID1 PLAVB
AC Q00738; STANDARD; PRT; 2869 AA.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -|- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
human reticulocyte cells.
CC -|- SUBUNIT: Homodimer (Potential).
CC -|- SUBCELLULAR LOCATION: Membrane-bound.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; M88097; AAA29743.1; -
CC Malaria; Receptor; Signal; Transmembrane.
KW SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCEFF CRC64;

Query Match 3.6%; Score 149; DB 1; Length 2869;

Best Local Similarity 17.2%; Pred. No. 7.1;

Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

QY 24 KOATQKTENLTPDEVSKKEGINAEQ-----IVIKITD-----QGY---VTSH 62

Db 1702 KESYEKNLETQONEMSR---INVEEGSLTDIXKIITDIENDLLMKKKQYEEGLQKIKEN 1758

QY 63 GDHYHYTYNGKVPYDAIISEELMKDPN-----YKLDEDIVNEKGVGVKVDGKYVY 116

Db 1759 AD-----KRKSNFELVGSSEINALLDPSTSIKILKLEYDMTGLK-NYGVMEIHGEF 1812

QY 117 LKD-----AAHADN-----VRTKEINRQKQHSQHRE----- 144

Db 1813 TKSYNLIETHLSNATDYSVTFEKAQSLRELAEEHEHLRREBEAIFLLNDIKKVESLKL 1872

QY 145 -----GTPRNDGAVALARSGRYTTD-----GVFNASDIIDTG 181

Db 1873 LKEMMKVSAEYEGMKDHTSVSQLVQDMKTIIVDELKTLNDISECSSVNNVSVIVKVK 1932

QY 182 DAYIVPHGDHYHYTPKNELSASAAAAEFLSGRGNLSNRTYRQNSDNTS---RTNW 238

Db 1933 ES---KADYRR--DANSWYESWTLANYFLSLDEAKISSGMEFNAEMKSNFKTDLLEIF 1987

QY 239 PVSVPNTTNTNSNN-----TNSQASQSNIDSLKQLYKPLSQRHVESDGLV 290

Db 1988 SVISNELLKKIQDSNDVIQKERESEQLAKADATIVNVIK--LKNEFNEKLEAKNKE 2045

QY 291 FDPAQITSRARGVAVPHGDHYHYEIPYSQM---SELEERTARIIPLYRSNHWVPDSRPE 347

Db 2046 EVSEKVRALKRLSQVQEGIRCFHFNRLLDNTTELENLKKMVTI-YR-----DKSE 2098

QY 348 QPSPQTPPEPSPQAPNPKIDSNSIVSQLVRKGVGYVEEKGISRYVFAKDLPSSET 407

Db 2099 RES-----GLQEMEN-EMTYSNSITQL-----EGIVVSAGESKEDIKLERNEE 2143

QY 408 VKNLESKLSQESVSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGKNSDFQAL 467

Db 2144 MRNISEKIS-----TIDSKVIEMNSTIDELYKLGKN--COAHWISLSYTA--MKT 2191

QY 468 DKLLERLNDSTNKEKLVDDLLAFLAIPHTPERLGKNSQIETVEDEVRIALQADKYTTS 527

Db 2192 SKLLIMINKEKTEKCV-----YKONSST 2219

QY 528 DGYI-----FDEHDIISDEGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKE 575

Db 2220 DGYVETLKGFGYKSLTFSSASEIVQADTVSNFAGHE-----KESLNAIRDIKKELYLPH 2275

QY 576 KGILPSPDADVANKPTG-DSAAAIYNRVKGK-----IPLVRLPYVMEHTVEV-- 624

Db 2276 Q-----NSDISIVEGGVQNMALYDKLNEEKREMDLYRNISETKLMQM-EHSTDFVK 2327

QY 625 -----KNGNLIIP----- 632

Db 2328 PMIELHKGMMETNNKSLLEKELKLSVNDHMSMEAMIKNGLKYTPESVQNNIYSVI 2387

QY 633 -----HKDHY-----HNTKFAWFDHDT-----YKAPNYTLEDLFATI 665

Db 2388 EAEVKTLIEDRDYGDNYQIVVEHKKQFSLIDRTNALMDDIIFKKNENYLN----- 2440

666 KYVHEPDERHS-NDGWNASEHVLGKDHSEDPNKNFKADEPVBETPAEP----- 717
 2441 ---MEVNTETHRVNDYIEKTNKLVQAKTEYEQILENIKONDWMLQIFLKKSIIIFYF 2497
 718 EYPQVETKVEAQLAEVLLAKVTD--SSLKANATETLAGLRNLLTIQIMDNNSIMAEA 775
 2498 ENVKKKESILNDLYEQERLL-KIGEHLDEIKRNVETILSSVEIDOKMEMSKN-LLEKK 2555
 776 EKLL 779
 2556 SKMM 2559

RESULT 15
 CUT7 SCHPO STANDARD; PRT; 1085 AA.
 AC P24339;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-like protein cut7.
 GN CUT7 OR SPAC25G10.07C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91015362; PubMed=2145514;
 RA Hagan I., Yanagida M.
 RT "Novel potential mitotic motor protein encoded by the fission yeast
 cut7+ gene."
 RL Nature 347:563-566(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Waleh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkhardt G., Aert R., Robben J., Grymonprez B.,
 RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RL "The genome sequence of Schizosaccharomyces pombe."
 Nature 415:871-880(2002).
 CC -1- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM
 CC G2 TO M PHASE OF THE CELL CYCLE, THE SPINDLE POLE BODY DUPLICATES;
 CC THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO
 CC FORM A SHORT SPINDLE THAT ELONGATES TOWARD THE NUCLEUS AT
 CC METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. B1MC
 CC SUBFAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X57513; CAA40738.1; -;
 CC EMBL; Z70691; CAA94636.1; -;
 CC F1R; T38378; T38378.
 CC HSP; F17119; 3KAR.
 CC GeneDB SPOMBE; SPAC25G10.07c; -;
 CC InterPro; IPR001752; kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KISC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Cell division; Microtubule; ATP-binding; Coiled coil;
 KW Mitosis; Cell cycle; Phosphorylation; Repeat.
 FT DOMAIN 70 435
 FT KINESIN-MOTOR (BY SIMILARITY).
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 436 604
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 715 740
 FT COILED COIL (POTENTIAL).
 FT NP_BIND 159 166
 FT ATP (BY SIMILARITY).
 FT REPEAT 987 998
 FT REPEAT 999 1010
 FT MOD_RES 1011 1011
 FT PHOSPHORYLATION (BY CDC2) (BY
 FT SIMILARITY)
 FT SASNPKRREPTIDTGYDPRSDTNSPT ->
 FT LRAILGNDSVLLSLTL (IN REF. 1).
 FT SEQUENCE 1085 AA; 122133 MW; 5669277875559D58 CRC64;
 SQ
 Query Match 3 5%; Score 145.5; DB 1; Length 1085;
 Best Local Similarity 18.6%; Pred. No. 2.8; Indels 301; Gaps 41;
 Matches 182; Conservative 136; Mismatches 357;
 QY 23 GKQATQKTENLTPDEVSKREGINAE-----QIVIKI--TDQYVTSKGDHYHYN 70
 DB 159 GQTGCTKTYTMSGD-LSDSDGILSEGAGLIPRALYQLFSSLDNSNQYAVK-CSYVELYN 216
 QY 71 GKVPYDAITSELLMKDPNKLKDED-----IVNEKGGVIVKDGKYYVYLKDAHADN 125
 DB 217 EEI-RDLLVSEEL--RKPARVFDTSRRGNVVITGIEEY-INKAGDGLRLREGSHRRQ 272
 QY 126 VRTKEINQKQEHSS-----QHREGTPRNDGAVALARSGRYTTDDGVIFNASDIIEDT 180
 DB 273 VAATKCNLSSSHSIFTLHKKVSSGMTDETSLTINN-----NSDILLR-- 319
 QY 181 GDAYIVPHGDHYHYIPKNELASLAAAFILSGRGNLSNRT----- 223
 DB 320 -----ASKLHMVDLAGSENI--GRSGAENKRARETGMINSILLTLGRVI 361
 QY 224 -----YR-----RONSNTSRTNWVPSVNPPTNTNTNNSNTNSQASQS 264
 DB 362 NALVEKAHPIVRESKLTLLQDLSGKTKTSMITVVS---STNTLEETISTLEYAFA 418
 QY 265 NDIDSLKLYKLPLSRHVESDGLVDPDAQI-----TSRTARGVAVPHGDHYHYFIPYSQ 319
 DB 419 K---SIRNKPQNQLVFRKVLKDLVLDIERKNDLNATKNGVLAESTYKELMDRVQ 475
 QY 320 MSEL-----EERIAIIRPLRYRNHWPDRPEQSPQPTPEPSGPQAPNLKI----- 369
 DB 476 NKDLLCQEQARKLEVLDLNVKS-----SREQLYVYSKSNQEHKKEVEALQLQLVNSSTE 529
 QY 370 -----DSNSSLVSQLYRKVGEYVFEKSGISRVVFAKDLFSETVKNLESKLSKQESVSH 424
 DB 530 LESVKGSEKLNKELVLEIEKRYKNEAKITTVATDLQQYRESKEYIASLYEKLDFT 589
 QY 425 LTAKENVAPRDQEFYDKAYNLL-----TEAH 451
 DB 590 ERNNKEN----ENNFWNLKENLLTLRLSPHGSFTDETNGYFTLLDNFNASMEELLNTHSN 645